# Two quantum analogues of Fisher information from a large deviation viewpoint of quantum estimation

**Masahito Hayashi†**

† Laboratory for Mathematical Neuroscience, Brain Science Institute, RIKEN
2-1, Hirosawa, Wako, Saitama 351-0198, Japan. e-mail masahito@brain.riken.go.jp

**Abstract.** We discuss two quantum analogues of the Fisher information, the symmetric logarithmic derivative (SLD) Fisher information and Kubo-Mori-Bogoljubov (KMB) Fisher information from a large deviation viewpoint of quantum estimation and prove that the former gives the true bound and the latter gives the bound of superefficient estimators. It is shown that the difference between them is characterized by the change of the order of limits.

Submitted to: *J. Phys. A: Math. Gen.*

PACS numbers: 03.67.-a,02.50.Tt

## 1. Introduction

As is well known, there is serious ambiguity concerning the order among non-commutative operators in the quantization of products of several variables. A similar ambiguity is observed in a quantum extension of Fisher information, which plays an important role in the parameter estimation for a probability distribution family and is, in a sense, the unique inner product satisfying invariance. Indeed, its quantum version cannot be uniquely determined for a quantum state family, and their geometrical properties has been discussed by many authors. One is the Kubo-Mori-Bogoljubov (KMB) Fisher inner product $\tilde{J}_\rho$, which can be regarded as the canonical correlation from the viewpoint of the linear response theory in statistical mechanics. It is defined by

$$\tilde{J}_\rho(A,B) := \operatorname{Tr} A^* \tilde{L}_B, \quad \int_0^1 \rho^t \tilde{L}_B \rho^{1-t}\, d = B, \tag{1}$$

where $\rho$ is the density operator, and $A$ and $B$ are operators. It can be characterized as the limit of quantum relative entropy, which plays an important role in several topics of quantum information theory, for example, quantum channel coding [1][2], quantum source coding [3][4][5] and quantum hypothesis testing [6][7]. Another is the symmetric logarithmic derivative (SLD) Fisher inner product

$$J_\rho(A,B) := \operatorname{Tr} A^* L_B, \quad \frac{1}{2}(L_B\rho + L_B\rho) = B, \tag{2}$$

which is closely related to the achievable lower bound of mean square error (MSE) not only for the one-parameter case [8][9][10], but also for the multi-parameter case [11][12][13] in quantum estimation. Its difference can be regarded as the difference in the order of the operators, and is influenced by the many ways of defining Fisher information for a probability distribution family.

However, there are a few previous papers in which both are discussed from a unified viewpoint. In this paper, to clarify the difference in a unified context, we introduce the large deviation viewpoint of quantum estimation as a unified viewpoint. This type of comparison was initiated by Nagaoka [14]. In this paper, we discuss this type of comparison more deeply. To summarize the main results, we need to summarize the classical estimation theory including Bahadur's large deviation theory in section 2. After this summary, we can briefly outline the main results in section 3, i.e., the difference is unitedly characterized from three contexts. To simplify the notations, even if we need the Gauss notation [ ], we omit it when we do not seem to confuse. Some proofs are very complicated, and we present them in the Appendix.

## 2. Summary of classical estimation theory

We summarize the relationship between the parameter estimation for the probability distribution family $\{p_\theta | \theta \in \Theta \subset \mathbb{R}\}$ and its Fisher information. One of the definitions of Fisher information is given by

$$J_\theta := \int_\Omega l_\theta(\omega)^2 p_\theta(\omega)\, d\omega, \tag{3}$$

where the logarithmic derivative $l_\theta(\omega)$ is defined as

$$l_\theta(\omega) := \frac{\frac{dp_\theta(\omega)}{d\theta}}{p_\theta(\omega)}. \tag{4}$$

Using the relative entropy (Kullback-Leibler divergence) $D(p\|q) := \int_\Omega (\log p(\omega) - \log q(\omega))\, p(\omega)\, d\omega$, we can define the Fisher information by another way:

$$J_\theta := \lim_{\epsilon \to 0} \frac{2}{\epsilon^2} D(p_{\theta+\epsilon}\|p_\theta). \tag{5}$$

These two definitions coincide under some regularity conditions for a family. For an estimator that is defined as a map from the data set $\Omega$ to the parameter set $\Theta$, we sometimes consider the *unbiasedness* condition:

$$\int_\Omega T(\omega) p_\theta(\omega)\, d\omega = \theta, \quad \forall \theta \in \Theta. \tag{6}$$

Using Schwartz inequality w.r.t. the inner product $\langle X, Y\rangle := \int_\Omega X(\omega) Y(\omega) p_\theta(\omega)\, d\omega$ for variables $X, Y$, we can prove the following inequality (Cramér-Rao inequality) for any unbiased estimator $T$:

$$\int_\Omega (T(\omega) - \theta)^2 p_\theta(\omega)\, d\omega \geq \frac{1}{J_\theta}. \tag{7}$$

When the number of data $\vec{\omega}_n := (\omega_1, \ldots, \omega_n)$, which obeys the unknown probability $p_\theta$, is sufficiently large, we discuss a sequence $\{T_n\}$ of estimators $T_n(\vec{\omega}_n)$. If $\{T_n\}$ is suitable as a sequence of estimators, we can expect that it converges to the true parameter $\theta$ in probability, i.e., it satisfies the *weakly consistent* condition:

$$\lim_{n\to\infty} p_\theta\{|T_n - \theta| > \epsilon\} = 0, \quad \forall \epsilon > 0, \forall \theta \in \Theta. \tag{8}$$

Usually, the performance of a sequence $\{T_n\}$ of estimators is measured by the speed of its convergence. As one criterion, we focus on the speed of the convergence in mean square error. If a sequence $\{T_n\}$ of estimators satisfies the weakly consistent condition and some regularity conditions, the asymptotic version of Cramér-Rao inequality:

$$\lim_{n\to\infty} n \int_\Omega (T_n(\vec{\omega}_n) - \theta)^2 p_\theta^n(\omega)\, d\omega \geq \frac{1}{J_\theta} \tag{9}$$

holds. If it satisfies only the weakly consistent condition, it is possible that it surpasses the bound of (9) at only one point. Such a sequence of estimators is called superefficient. We can reduce its error to any amount at one point under the weakly consistent condition (8)

As another criterion, we evaluate the decreasing rate of the tail probability:

$$\beta(\{T_n\}, \theta, \epsilon) := \lim_{n\to\infty} \frac{-1}{n} \log p_\theta^n\{|T_n - \theta| > \epsilon\}. \tag{10}$$

This method was initiated by Bahadur [15]. From the monotonicity of the divergence, we can prove the inequality

$$\beta(\{T_n\}, \theta, \epsilon) \leq \min\{D(p_{\theta+\epsilon}\|p_\theta), D(p_{\theta-\epsilon}\|p_\theta)\} \tag{11}$$

for any weakly consistent sequence $\{T_n\}$ of estimators. Its proof is essentially given in our proof of Theorem 2. Since it is difficult to analyze $\beta(\{T_n\},\theta,\epsilon)$ except for an exponential family, we focus on another quantity $\alpha(\{T_n\},\theta) := \lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\beta(\{T_n\},\theta,\epsilon)$. For the exponential family case, see Appendix H. From the inequality (11), the inequality

$$\alpha(\{T_n\},\theta) \le \frac{J_\theta}{2} \tag{12}$$

holds. If $T_n$ is the maximum likelihood estimator (MLE), the equality of (12) holds under some regularity conditions for the family [15] [16]. This type of discussion is different from the MSE type of discussion in deriving (12) only from the weakly consistent condition. Therefore, there is no superefficient estimator w.r.t. the large deviation evaluation.

Next, we summarize the relationship between Stein's lemma in the hypothesis testing and the above large deviation type of discussion in the estimation. In the hypothesis testing, we decide that the null hypothesis be accepted or rejected from the data $\vec{\omega}_n := (\omega_1,\ldots,\omega_n)$ which obeys an unknown probability. For the decision, we need to define an *accept region* $A_n$ as a subset of $\Omega^n$. If the null hypothesis is $p$ and the alternative is $q$, the first error (though the true distribution is $p$, we reject the null hypothesis.) probability $\beta_{1,n}(A_n)$ and the second error (though the true distribution is $q$, we accept the null hypothesis.) probability $\beta_{2,n}(A_n)$ are given by

$$\beta_{1,n}(A_n) := 1 - p^n(A_n), \quad \beta_{2,n}(A_n) := q^n(A_n).$$

Regarding the decreasing rate of the second error probability under the constant constraint of the first error probability, the equation

$$\lim_{n\to\infty}\frac{-1}{n}\log\min\{\beta_{2,n}(A_n)|\beta_{1,n}(A_n)\le\epsilon\} = D(p\|q), \quad \epsilon > 0 \tag{13}$$

holds (Stein's lemma). The inequality (11) can be derived from this lemma. We can regard the large deviation type of evaluation in the estimation to be the limit of Stein's lemma.

## 3. Outline of main results

Let us return to the quantum case. We can define several quantum analogues of Fisher information following the choice of the quantum analogue of logarithmic derivative. First, we define the symmetric logarithmic derivative (SLD) $L_\theta$ for a state family $\{\rho_\theta\in\mathcal{S}(\mathcal{H})|\theta\in\Theta\}$ by

$$\frac{d\rho_\theta}{d\theta} = \frac{1}{2}(L_\theta\rho_\theta + \rho_\theta L_\theta). \tag{14}$$

Following to (14) and (3), we can define the SLD Fisher information $J_\theta$ as

$$J_\theta := \mathrm{Tr}\, L_\theta^2\rho_\theta, \tag{15}$$

which corresponds to the SLD Fisher inner product. If the state $\rho_\theta$ is nondegenerate, SLD $L_\theta$ is not uniquely determined. However, as is proven in Appendix A, the SLD

Fisher information $J_\theta$ is uniquely determined, i.e., it is independent of the choice of the SLD $L_\theta$. When we regard it as an inner product, it is the minimum one among invariant inner products [17].

Next, as another quantum analogue of logarithmic derivative, we define the Kubo-Mori-Bogoljubov (KMB) logarithmic derivative $\tilde{L}_\theta$ as

$$\int_0^1 \rho_\theta^t \tilde{L}_\theta \rho_\theta^{1-t}\, dt = \frac{d\rho_\theta}{d\theta}.$$

Note that $\tilde{L}_\theta$ has another form

$$\tilde{L}_\theta = \frac{d \log \rho_\theta}{d\theta}.$$

Therefore, we can define the KMB Fisher information $\tilde{J}_\theta$ as

$$\tilde{J}_\theta = \int_0^1 \mathrm{Tr}\, \rho_\theta^t \tilde{L}_\theta \rho_\theta^{1-t} \tilde{L}_\theta\, dt,$$

which corresponds to the KMB Fisher inner product $\tilde{J}_\rho$. This inner product can be regarded as the canonical correlation from the viewpoint of the linear response theory in statistical physics. (See Chap. 7 in Amari and Nagaoka [18], Petz and Toth [19], Petz[17] or Petz and Sudár [20].) If we follow the second definition (5), the KMB Fisher information $\tilde{J}_\theta$ is suitable because the equation

$$\tilde{J}_\theta = \lim_{\epsilon \to 0} \frac{2}{\epsilon^2} D(\rho_{\theta+\epsilon} \| \rho_\theta) \tag{16}$$

holds, where $D(\rho\|\sigma)$ is the quantum relative entropy $\mathrm{Tr}\, \rho(\log \rho - \log \sigma)$.

As another quantum analogue, the RLD Fisher information $\check{J}_\theta$:

$$\check{J}_\theta := \mathrm{Tr}\, \rho_\theta \check{L}_\theta \check{L}_\theta^*, \quad \frac{d\rho_\theta}{d\theta} = \rho_\theta \check{L}_\theta$$

is well known. When we regard it as an inner product, it is the maximum one among invariant inner products [17]. Since it is not useful in the one-parameter case, we do not discuss it, for the meantime. Since the difference in definition can be regarded as the difference in the order of operators, these quantum analogues coincide when all states of the family are commutative with each other. However, in the general case, they do not coincide and the inequality $\tilde{J}_\theta \geq J_\theta$ holds, as exemplified in section 4.

In the following, we consider the roles these quantum analogues of Fisher information play in the parameter estimation for the state family. As is discussed in detail in section 4, the estimator is described by the pair of POVM $M$ (which corresponds to the measurement and is defined in section 4) and the map from the data set to the parameter space $\Theta$. Similarly to the classical case, we can define an unbiased estimator. For any unbiased estimator $E$, the SLD Cramér-Rao inequality

$$V(E) \geq \frac{1}{J_\theta} \tag{17}$$

holds, where $V(E)$ is the mean square error (MSE) of the estimator $E$.

Next, we consider an asymptotic setting. As a quantum analogue of the $n$-i.i.d. condition, we treat the quantum $n$-i.i.d. condition, i.e., we consider the case that the number of systems that are independently prepared in the same unknown state is sufficiently large in section 5. In this case, the measurement is denoted by a POVM $M^n$ on the composite system $\mathcal{H}^{\otimes n}$ and the state is described by the density $\rho^{\otimes n}$. Of course, such POVMs include a POVM that requires quantum correlations between the respective quantum systems. Similarly to the classical case, for a sequence $\vec{E} = \{E^n\}$ of estimators, we can define the *weakly consistent* condition given in (30). Regarding the large deviation type of evaluation, as is discussed in section 5, we can similarly define the quantities $\beta(\vec{E},\theta,\epsilon), \alpha(\vec{E},\theta)$. Similarly to (11)(12), under the weakly consistent (WC) condition, the inequalities

$$\begin{aligned}\beta(\vec{E},\theta,\epsilon) &\le \min\{D(\rho_{\theta+\epsilon}\|\rho_\theta), D(\rho_{\theta-\epsilon}\|\rho_\theta)\}\\ \alpha(\vec{E},\theta) &\le \frac{1}{2}\tilde{J}_\theta\end{aligned}$$

hold. From these discussions, the bound in the large deviation type of evaluation seems different from the one in the MSE case. However, as mentioned in section 6, roughly speaking, the inequality

$$\alpha(\vec{E},\theta) \le \frac{1}{2}J_\theta \tag{18}$$

holds if the sequence $\vec{E}$ satisfies the strongly consistent (SC) condition introduced in section 6 as a stronger condition. As is mentioned in section 7, these bounds can be attained in their respective senses. Therefore, roughly speaking, the difference between the two quantum analogues can be regarded as the difference of consistent conditions and can be characterized as

$$\begin{aligned}\sup_{\vec{E}:\mathrm{SC}} \lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\beta(\vec{E},\theta,\epsilon) &= \frac{1}{2}J_\theta\\ \sup_{\vec{E}:\mathrm{WC}} \lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\beta(\vec{E},\theta,\epsilon) &= \frac{1}{2}\tilde{J}_\theta.\end{aligned}$$

Even if we restrict our estimators to strongly consistent ones, the difference in both appears as

$$\sup_{\vec{M}:\mathrm{SC}} \liminf_{\epsilon\to 0}\frac{1}{\epsilon^2}\beta(\vec{M},\theta,\epsilon) = \frac{J_\theta}{2} \tag{19}$$

$$\liminf_{\epsilon\to 0}\frac{1}{\epsilon^2}\sup_{\vec{M}:\mathrm{SC}}\beta(\vec{M},\theta,\epsilon) = \frac{\tilde{J}_\theta}{2}, \tag{20}$$

where, for a precise statement, as expressed in section 9, we need more complicated definitions.

However, we should think that, the real bound is the bound $\frac{J_\theta}{2}$ for the following two reasons. The first reason is the fact that we can construct the sequence of estimators attaining the bound $\frac{J_\theta}{2}$ at all points, which is proven in section 7. On the other hand, there is a sequence of estimators attaining the bound $\frac{\tilde{J}_\theta}{2}$, but it cannot attain the bound at all points. The other reason is the naturalness of the conditions for deriving the bound $\frac{J_\theta}{2}$. In other words, an estimator attaining $\frac{J_\theta}{2}$ is natural, but an estimator attaining $\frac{\tilde{J}_\theta}{2}$ is very irregular. Such a sequence of estimators can be regarded as a superefficient estimator and does not satisfy other regularity conditions than the weakly consistent condition. This type of discussion of the superefficiency is different from the MSE type of discussion in that any superefficient estimator is bounded by the inequality (18).

To consider the difference between the two quantum analogues of the Fisher information more deeply, we need to analyze how to achieve the bound $\frac{\tilde{J}_\theta}{2}$. It is important for this analysis to consider the relationship between the above discussion and the quantum version of Stein's lemma. Similarly to the classical case, when the null hypothesis is the state $\rho$ and the alternative is the state $\sigma$, we evaluate the decreasing rate of the second error probability under the constant $\epsilon$ constraint of the first error probability. It is well known as quantum Stein's lemma, in which its exponential component is given by the quantum relative entropy $D(\rho\|\sigma)$ for any $\epsilon$. Hiai and Petz [6] constructed a sequence of tests to attain the optimal rate $D(\rho\|\sigma)$, by constructing the sequence $\{M^n\}$ of POVMs such that

$$\lim_{n\to\infty} \frac{1}{n} D(\mathrm{P}_\rho^{M^n} \| \mathrm{P}_\sigma^{M^n}) = D(\rho\|\sigma). \tag{21}$$

Ogawa and Nagaoka [7] proved that there is no test excelling the bound $D(\rho\|\sigma)$. It is known that by using the group representation theory, we can construct the POVM satisfying (21) independently of $\rho$ [21]. For the reader's convenience, we give a summary of this in Appendix G. As discussed in section 7.2, this type of construction is useful for the construction of an estimator attaining the bound $\frac{\tilde{J}_\theta}{2}$ at one point. Since the proper bound of the large deviation is $\frac{J_\theta}{2}$, we cannot regard the quantum estimation as the limit of quantum Stein's lemma.

In order to consider the properties of estimators attaining the bound $\frac{\tilde{J}_\theta}{2}$ at one point from another viewpoint, we consider the restriction that makes such a construction impossible. We introduce a class of estimators whose POVMs do not need a quantum correlation in section 8. In this class, we assume that the POVM on the $l$-th system is chosen from $l-1$ data. We call such an estimator an adaptive estimator. When an adaptive estimator $\vec{E}$ satisfies the weakly consistent condition, the inequality

$$\alpha(\vec{E},\theta) \le \frac{1}{2} J_\theta \tag{22}$$

holds (See section 6). In this class, we do not use quantum correlations. Similarly, we can define a class of estimators that use quantum correlations up to $m$ systems. We call such an estimator an $m$-adaptive estimator. For any $m$-adaptive weakly consistent estimator $\vec{E}$, inequality (22) holds. Therefore, it is impossible to construct a sequence of estimators attaining the bound $\frac{\tilde{J}_\theta}{2}$ if we fix the number of systems in which we use quantum correlations. As mentioned in section 8, taking limit $m \to \infty$, we have

$$\lim_{m\to\infty} \lim_{\epsilon\to 0} \sup_{\vec{M}: m\text{-AWC}} \frac{1}{\epsilon^2} \beta(\vec{M},\theta,\epsilon) = \frac{J_\theta}{2}, \tag{23}$$

where $m$-AWC denotes $m$-adaptive weakly consistent. However, as the third characterization of the difference between the two quantum analogues, as precisely mentioned in section 9, we have the following equation

$$\lim_{\epsilon\to 0}\lim_{m\to\infty}\sup_{\vec{M}:m\text{-ASC}}\frac{1}{\epsilon^2}\beta(\vec{M},\theta,\epsilon)=\frac{\tilde{J}_\theta}{2}, \tag{24}$$

where $m$-ASC denotes $m$-adaptive strongly consistent. Equation (24) is a stronger condition than (20). Equations (23) and (24) indicate that the order of limits $\lim_{m\to\infty}$ and $\lim_{\epsilon\to 0}$ is more crucial than the difference between two types of consistencies.

## 4. Summary of non-asymptotic setting in quantum estimation

In a quantum system, in order to discuss the probability distribution which the data obeys, we need to define a positive operator valued measure (POVM).

A POVM $M$ is defined as a map from Borel sets of the data set $\Omega$ to the set of bounded, self-adjoint and positive semi-definite operators, which satisfies

$$M(\emptyset)=0,\quad M(\Omega)=\mathrm{I},\quad \sum_i M(B_i)=M(\cup B_i)\text{ for disjoint sets.}$$

If the state on the quantum system $\mathcal{H}$ is a density operator $\rho$ and we perform a measurement corresponding to a POVM $M$ on the system, the data obeys the probability distribution $\mathrm{P}^M_\rho(B):=\mathrm{Tr}\,\rho M(B)$. If a POVM $M$ satisfies $M(B)^2=M(B)$ for any Borel set $B$, $M$ is called a projection-valued measure (PVM). The spectral measure of a self-adjoint operator $X$ is a PVM, and is denoted by $E(X)$. For $1>\lambda>0$ and any POVMs $M_1$ and $M_2$ taking values in $\Omega$, the POVM $B\mapsto\lambda M_1(B)+(1-\lambda)M_2(B)$ is called the *random combination* of $M_1$ and $M_2$ in the ratio $\lambda:1-\lambda$. Even if $M_1$'s data set $\Omega_1$ is different from $M_2$'s data set $\Omega_2$, $M_1$ and $M_2$ can be regarded as POVMs taking values in the disjoint union set $\Omega_1\coprod\Omega_2:=(\Omega_1\times\{1\})\cup(\Omega_2\times\{2\})$. In this case, we can define a random combination of $M_1$ and $M_2$ as a POVM taking values in $\Omega_1\coprod\Omega_2$ and call it the *disjoint random combination*. In this paper, we simplify the probability $\mathrm{P}^M_{\rho_\theta}$ and the relative entropies $D(\rho_{\theta_0}\|\rho_{\theta_1}), D(\mathrm{P}^M_{\rho_{\theta_0}}\|\mathrm{P}^M_{\rho_{\theta_1}})$ to $\mathrm{P}^M_\theta$, $D(\theta_0\|\theta_1)$ and $D^M(\theta_0\|\theta_1)$, respectively.

In the one-parameter quantum estimation, the estimator is described by a pair comprising a POVM and a map from its data set to the real number set $\mathbb{R}$. Since the POVM $M\circ T^{-1}$ takes values in the real number set $\mathbb{R}$, we can regard any estimator as a POVM taking values in the real number set $\mathbb{R}$. In order to evaluate MSE, Helstrom [8, 9] derived the SLD Cramér-Rao inequality as a quantum counterpart of Cramér-Rao inequality (28). If an estimator $M$ satisfies that

$$\int_{\mathbb{R}} x\,\mathrm{Tr}\,\rho_\theta M(\,dx)=\theta,\quad \forall\theta\in\Theta, \tag{25}$$

it is called unbiased. If $\theta-\theta_0$ is sufficiently small, we can obtain the following approximation in the neighborhood of $\theta_0$:

$$\int_{\mathbb{R}} x\,\mathrm{Tr}\,\rho_{\theta_0}M(\,dx)+\left(\int_{\mathbb{R}} x\,\mathrm{Tr}\left.\frac{\partial\rho_\theta}{\partial\theta}\right|_{\theta=\theta_0}M(\,dx)\right)(\theta-\theta_0)\cong\theta_0+(\theta-\theta_0).$$

It implies the following two conditions:

$$\int_{\mathbb{R}} x \operatorname{Tr} \left.\frac{\partial \rho_\theta}{\partial \theta}\right|_{\theta=\theta_0} M(\,dx) = 1 \tag{26}$$

$$\int_{\mathbb{R}} x \operatorname{Tr} \rho_{\theta_0} M(\,dx) \qquad = \theta_0. \tag{27}$$

If an estimator $M$ satisfies (26) and (27), it is called locally unbiased at $\theta_0$. For any locally unbiased estimator $M$ (at $\theta$), the inequality, which is called the SLD Cramér-Rao inequality,

$$\int_{\mathbb{R}} (x-\theta)^2 \operatorname{Tr} \rho_\theta M(\,dx) \geq \frac{1}{J_\theta} \tag{28}$$

holds. This inequality is derived from Schwartz inequality with respect to the SLD Fisher information $\langle X|Y\rangle := \operatorname{Tr} \rho_\theta \frac{XY+YX}{2}$ [8] [9] [10].

The equality of (28) holds when the estimator is given by the spectral decomposition $E(\frac{L_\theta}{J_\theta}+\theta)$ of $\frac{L_\theta}{J_\theta}+\theta$, where $L_\theta$ is the SLD at $\theta$ and is defined in (14). This implies that the SLD Fisher information $J_{\theta_0}$ is consistent with the Fisher information at $\theta_0$ of the probability family $\left\{ \mathrm{P}_\theta^{E(\frac{L_{\theta_0}}{J_{\theta_0}}+\theta_0)} \middle| \theta \in \Theta \right\}$. The monotonicity of quantum relative entropy [22] [23] gives the following evaluation of the probability family $\left\{ \mathrm{P}_\theta^{E(\frac{L_{\theta_0}}{J_{\theta_0}}+\theta_0)} \middle| \theta \in \Theta \right\}$ as:

$$D^{E\left(\frac{L_{\theta_0}}{J_{\theta_0}}+\theta_0\right)}(\theta\|\theta_0) \leq D(\theta\|\theta_0).$$

Taking the limit $\theta \to \theta_0$, we have

$$J_\theta \leq \tilde{J}_\theta. \tag{29}$$

In this paper, we discuss the inequality (29) from the viewpoint of the large deviation type of evaluation of the quantum estimation. As simple examples of the one-parameter quantum state family, the following are known.

**Example 1 [One-parameter equatorial spin 1/2 system state family]**:

$$\mathcal{S}_r := \left\{ \rho_\theta := \frac{1}{2} \begin{pmatrix} 1+r\cos\theta & r\sin\theta \\ r\sin\theta & 1-r\cos\theta \end{pmatrix} \middle| 0 \leq \theta < 2\pi \right\}$$

In this family, we calculate

$$D(\rho_\theta\|\rho_0) = \frac{r}{2}(1-\cos\theta)\log\frac{1+r}{1-r}$$

$$\tilde{J}_\theta = \frac{r}{2}\log\frac{1+r}{1-r}$$

$$J_\theta = r^2.$$

Since the relations $\tilde{J}_\theta = \infty$, $J_\theta = 1$ hold in the case of $r = 1$, the two quantum analogues are completely different.

**Example 2 [One-parameter quantum Gaussian state family and half-line quantum Gaussian state family]**: We define the boson coherent vector $|\alpha\rangle := e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle$, where $|n\rangle$ is the number vector on $L^2(\mathbb{R})$. The quantum Gaussian state is defined as:

$$\rho_\theta := \frac{1}{\pi \overline{N}} \int_{\mathbb{C}} |\alpha\rangle\langle\alpha| e^{-\frac{|\alpha-\theta|^2}{\overline{N}}} \, d^2\alpha, \quad \forall \theta \in \mathbb{C}.$$

We call $\{\rho_\theta | \theta \in \mathbb{R}\}$ the one-parameter quantum Gaussian state family, and call $\{\rho_\theta | \theta \geq 0 (\theta \in \mathbb{R}^+ = [0, \infty))\}$ the half-line quantum Gaussian state family. In this family, we can calculate

$$D(\rho_\theta \| \rho_{\theta_0}) = \log\left(1 + \frac{1}{\overline{N}}\right) |\theta - \theta_0|^2,$$

$$\tilde{J}_\theta = 2\log\left(1 + \frac{1}{\overline{N}}\right),$$

$$J_\theta = \frac{2}{\overline{N} + \frac{1}{2}}.$$

## 5. The bound under the weakly consistent condition

We introduce the quantum independent-identical density (i.i.d.) condition in order to treat an asymptotic setting. Suppose that $n$-independent physical systems are given in the same state $\rho$. Then, the quantum state of the composite system is described by

$$\rho^{\otimes n} := \underbrace{\rho \otimes \cdots \otimes \rho}_{n} \text{ on } \mathcal{H}^{\otimes n},$$

where the tensored space $\mathcal{H}^{\otimes n}$ is defined by

$$\mathcal{H}^{\otimes n} := \underbrace{\mathcal{H} \otimes \cdots \otimes \mathcal{H}}_{n}.$$

We call this condition the quantum i.i.d. condition, which is a quantum analogue of the independent-identical distribution condition. In this setting, any estimator is described by a POVM $M^n$ on $\mathcal{H}^{\otimes n}$, whose data set is $\mathbb{R}$. In this paper, we simplify $\mathrm{P}^{M^n}_{\rho_\theta^{\otimes n}}$ and $D(\mathrm{P}^{M^n}_{\rho_{\theta_0}^{\otimes n}} \| \mathrm{P}^{M^n}_{\rho_{\theta_1}^{\otimes n}})$ to $\mathrm{P}^{M^n}_\theta$ and $D^{M^n}(\theta_0 \| \theta_1)$. The notation $M \times n$ denotes the POVM in which we perform the POVM $M$ to the respective $n$ systems.

**Definition 1 [Weakly consistent condition]**: A sequence of estimators $\vec{M} := \{M^n\}_{n=1}^{\infty}$ is called *weakly consistent* if

$$\lim_{n\to\infty} \mathrm{P}^{M^n}_\theta \left\{ |\hat{\theta} - \theta| > \epsilon \right\} = 0, \quad \forall \theta \in \Theta, \forall \epsilon > 0, \tag{30}$$

where $\hat{\theta}$ is the estimated value.

Now, we focus on the exponential component of the tail probability as follows:

$$\beta(\vec{M}, \theta, \epsilon) := \limsup_{n\to\infty} \frac{-1}{n} \log \mathrm{P}^{M^n}_\theta \left\{ |\hat{\theta} - \theta| > \epsilon \right\}.$$

We usually discuss the following value in stead of $\beta(\vec{M},\theta,\epsilon)$

$$\alpha(\vec{M},\theta) := \limsup_{\epsilon\to 0} \frac{1}{\epsilon^2}\beta(\vec{M},\theta,\epsilon) \tag{31}$$

because it is too difficult to discuss $\beta(\vec{M},\theta,\epsilon)$. The following theorem can be proven from the monotonicity of the quantum relative entropy.

**Theorem 2** *If a POVM $M^n$ on $\mathcal{H}^{\otimes n}$ satisfies the weakly consistent condition (30), the inequalities*

$$\beta(\vec{M},\theta,\epsilon) \le \inf\{D(\rho_{\theta'}\|\rho_\theta) | |\theta-\theta'| < \epsilon\} \tag{32}$$

$$\alpha(\vec{M},\theta) \quad \le \frac{\tilde{J}_\theta}{2} \tag{33}$$

*hold.*

Even if the parameter set $\Theta$ is not open (ex. the closed half-line $\mathbb{R}^+ := [0,\infty)$), this theorem holds.

*Proof:* The monotonicity of the quantum relative entropy yields that

$$D(\rho_{\theta'}^{\otimes n}\|\rho_\theta^{\otimes n}) \ge p_{n,\theta'}\log\frac{p_{n,\theta'}}{p_{n,\theta}} + (1-p_{n,\theta'})\log\frac{1-p_{n,\theta'}}{1-p_{n,\theta}},$$

for any $\theta'$ satisfying $|\theta'-\theta| > \epsilon$, where we denote the probability $\mathrm{P}_{\theta''}^{M^n}\left\{|\hat{\theta}-\theta| > \epsilon\right\}$ by $p_{n,\theta''}$. Using the inequality $-(1-p_{n,\theta'})\log(1-p_{n,\theta}) \ge 0$, we have

$$-\frac{\log \mathrm{P}_\theta^{M^n}\left\{|\hat{\theta}-\theta| > \epsilon\right\}}{n} = -\frac{\log p_{n,\theta}}{n} \le \frac{D(\rho_{\theta'}^{\otimes n}\|\rho_\theta^{\otimes n}) + h(p_{n,\theta'})}{np_{n,\theta'}}, \tag{34}$$

where $h$ is the binary entropy defined by $h(x) := -x\log x - (1-x)\log(1-x)$. Since the assumption guarantees that $p_{n,\theta'} \to 1$, the inequality

$$\beta(\vec{M},\theta,\epsilon) \le D(\rho_{\theta'}\|\rho_\theta) \tag{35}$$

holds, where we use the additivity of quantum relative entropy:

$$D(\rho_{\theta'}^{\otimes n}\|\rho_\theta^{\otimes n}) = nD(\rho_{\theta'}\|\rho_\theta).$$

Thus, we obtain (32). Taking the limit $\epsilon \to 0$ in inequality (35), we obtain (33). ■

As another proof, we can prove this inequality to be a corollary of quantum Stein's lemma [6, 7].

## 6. The bound under the strongly consistent condition

As discussed in section 4, the SLD Cramér-Rao inequality guarantees that the lower bound of MSE is given by the SLD Fisher information. Therefore, it is expected that the bound is connected with the SLD Fisher information for large deviation. In order to discuss the relationship between the SLD Fisher information and the bound for large deviation, we need another characterization with respect to the limit of the tail probability. We thus define

$$\begin{aligned}\underline{\beta}(\vec{M},\theta,\epsilon) &:= \liminf_{n\to\infty} \frac{-1}{n} \log \mathrm{P}_\theta^{M^n} \left\{ |\hat{\theta}-\theta| > \epsilon \right\} \\ \underline{\alpha}(\vec{M},\theta) &:= \liminf_{\epsilon\to 0} \frac{1}{\epsilon^2} \underline{\beta}(\vec{M},\theta,\epsilon). \end{aligned} \tag{36}$$

In the following, we try to link the quantity $\underline{\alpha}(\vec{M},\theta)$ with the SLD Fisher information. For this purpose, it is suitable to focus on an information quantity that satisfies the additivity and the monotonicity as in the proof of Theorem 1. Its limit should be the SLD Fisher information. The Bures distance $b(\rho,\sigma) := \sqrt{2(1-\mathrm{Tr}\,|\sqrt{\rho}\sqrt{\sigma}|)} = \sqrt{\min_{U:unitary} \mathrm{Tr}(\sqrt{\rho}-\sqrt{\sigma}U)(\sqrt{\rho}-\sqrt{\sigma}U)^*}$ is known to be an information quantity whose limit is the SLD Fisher information, as mentioned in Lemma 3. Of course, it can be regarded as a quantum analogue of Hellinger distance, and satisfies the monotonicity.

**Lemma 3** *[Uhlmann [24], Matsumoto [25]] If there exists an SLD $L_\theta$ satisfying (14), then the equation*

$$\frac{1}{4} J_\theta = \lim_{\epsilon\to 0} \frac{b^2(\rho_\theta,\rho_{\theta+\epsilon})}{\epsilon^2} \tag{37}$$

*holds.*

A proof of Lemma 3 is given in Appendix A. As discussed in the latter, Bures distance satisfies the monotonicity. Unfortunely, Bures distance does not satisfy the additivity.

However, the quantum affinity $I(\rho\|\sigma) := -8\log \mathrm{Tr}\left|\sqrt{\rho}\sqrt{\sigma}\right| = -8\log\left(1-\frac{1}{2}b(\rho,\sigma)^2\right)$ satisfies the additivity:

$$I(\rho^{\otimes n}\|\sigma^{\otimes n}) = nI(\rho\|\sigma). \tag{38}$$

Its classical version is called affinity and is introduced by Akahira and Takeuchi [26] in the following form:

$$I(p\|q) = -8\log\left(\sum_i \sqrt{p_i}\sqrt{q_i}\right). \tag{39}$$

As a trivial deformation of (37), the equation

$$\lim_{\epsilon\to 0} \frac{I(\rho_\theta\|\rho_{\theta+\epsilon})}{\epsilon^2} = J_\theta \tag{40}$$

holds. The quantum affinity satisfies the monotonicity w.r.t. any measurement $M$ (Jozsa [27], Fuchs [28]):

$$I(\rho\|\sigma) \ge I\left(\mathrm{P}_\rho^M \middle\| \mathrm{P}_\sigma^M\right) = -8\log\sum_\omega \sqrt{\mathrm{P}_\rho^M(\omega)}\sqrt{\mathrm{P}_\sigma^M(\omega)}. \tag{41}$$

The most simple proof of (41) is given in Fuchs [28] where it was directly proven that

$$\mathrm{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \le \sum_\omega \sqrt{\mathrm{P}_\rho^M(\omega)}\sqrt{\mathrm{P}_\sigma^M(\omega)}. \tag{42}$$

For the reader's convenience, a proof of (42) is given in Appendix B. From (38),(40) and (41), we can expect that the SLD Fisher information is in a sense closely related to a large deviation type of bound. From the additivity and the monotonicity of the quantum affinity, we can show the following lemma.

**Lemma 4** *The inequality*

$$4 \inf_{\{s|1\geq s\geq 0\}} \underline{\beta}'(\vec{M},\theta,s\delta)+\underline{\beta}'(\vec{M},\theta+\delta,(1-s)\delta)\leq I(\rho_\theta\|\rho_{\theta+\delta}) \tag{43}$$

*holds, where we define* $\underline{\beta}'(\vec{M},\theta,\delta):=\lim_{\epsilon\to+0}\underline{\beta}(\vec{M},\theta,\delta-\epsilon)$.

A proof of Lemma 4 is given in Appendix C. However, Lemma 4 cannot yield an inequality with respect to $\alpha(\vec{M},\theta)$ under the weakly consistent condition, as the inequality (35) does. Therefore, we consider a more strongly consistent condition than the weakly consistent condition (30).

**Definition 5 [Strongly consistent condition]**: A sequence of estimators $\vec{M} = \{M^n\}_{n=1}^{\infty}$ is called *strongly consistent* if the convergence of (36) is uniform for the parameter $\theta$ and if $\underline{\alpha}(\vec{M},\theta)$ is continuous for $\theta$. A sequence of estimators is called *strongly consistent at* $\theta$ if there exists a neighborhood $U$ of $\theta$ such that it is strongly consistent in $U$.

As a corollary of Lemma 4, we have the following theorem.

**Theorem 6** *Assume that there exists SLD* $L_\theta$ *satisfying (14). If a sequence of estimators* $\vec{M}=\{M^n\}_{n=1}^{\infty}$ *is strongly consistent at* $\theta$*, then the inequality*

$$\underline{\alpha}(\vec{M},\theta)\leq\frac{J_\theta}{2} \tag{44}$$

*holds.*

*Proof:* From the assumption, for any real $\epsilon>0$ and any element $\theta\in\Theta$, there exists a sufficiently small real $\delta>0$ such that $(\underline{\alpha}(\vec{M},\theta)-\epsilon)\epsilon'^2\leq\underline{\beta}'(\vec{M},\theta,\epsilon'),\underline{\beta}'(\vec{M},\theta+\delta,\epsilon')$ for $\forall\epsilon'<\delta$. Therefore, the inequality (43) yields the relations

$$\begin{aligned}2(\underline{\alpha}(\vec{M},\theta)-\epsilon)\delta^2&=4(\underline{\alpha}(\vec{M},\theta)-\epsilon)\inf_{\{s|1\geq s\geq 0\}}\left(s^2\delta^2+(1-s)^2\delta^2\right)\\ &\leq 4\inf_{\{s|1\geq s\geq 0\}}\underline{\beta}'(\vec{M},\theta,s\delta)+\underline{\beta}'(\vec{M},\theta+\delta,(1-s)\delta)\leq I(\rho_\theta\|\rho_{\theta+\delta}).\end{aligned} \tag{45}$$

Lemma 3 and (45) guarantee (44) for $\forall\theta\in\Theta$. ■

**Remark 1** Inequality (42) can be regarded as a special case of the monotonicity w.r.t. any trace-preserving CP (completely positive) map $C:\mathcal{S}(\mathcal{H}_1)\to\mathcal{S}(\mathcal{H}_2)$:

$$\left(\mathrm{Tr}\left|\sqrt{\rho}\sqrt{\sigma}\right|\right)^2\leq\left(\mathrm{Tr}\left|\sqrt{C(\rho)}\sqrt{C(\sigma)}\right|\right)^2. \tag{46}$$

which is proven by Jozsa [27] because the map $\rho\mapsto \mathrm{P}_\rho^M$ can be regarded as a trace-preserving CP map from the $C^*$ algebra of bounded operators on $\mathcal{H}$ to the commutative $C^*$ algebra $C(\Omega)$, where $\Omega$ is the data set.

## 7. Achievabilities of the bounds

Next, we discuss the achievabilities of the two bounds $\tilde{J}_\theta$ and $J_\theta$ in their respective senses. In this section, we discuss these achievabilities in two cases: the first case is the one-parameter quantum Gaussian state family, and the second case is an arbitrary one-parameter finite-dimensional quantum state family that satisfies some assumptions.

### *7.1. one-parameter quantum Gaussian state family*

In this subsection, we discuss these achievabilities in one-parameter quantum Gaussian state family.

**Theorem 7** *In the one-parameter quantum Gaussian state family, the sequence of estimators $\vec{M}^s = \{M^{s,n}\}_{n=1}^\infty$ (defined in the following) satisfies the strongly consistent condition and the relations*

$$\alpha(\vec{M}^s, \theta) = \underline{\alpha}(\vec{M}^s, \theta) = \frac{J_\theta}{2} = \frac{1}{\overline{N} + \frac{1}{2}}. \tag{47}$$

**[Construction of $\vec{M}^s$]**: We perform the POVM $E(Q)$ to all systems, where $Q$ is the position operator on $L^2(\mathbb{R})$. The estimated value $T_n$ is determined to be the mean value of $n$ data. ■

*Proof:* Since the equation

$$\mathrm{P}_{|\alpha\rangle\langle\alpha|}^{E(Q)}(\,dx) = \sqrt{\frac{2}{\pi}} e^{-2(x-\alpha_x)^2}\,dx$$

holds, we have the equation

$$\mathrm{P}_{\theta}^{E(Q)}(\,dx) = \sqrt{\frac{2}{\pi(2\overline{N}+1)}} e^{-\frac{2(x-\theta)^2}{2\overline{N}+1}}\,dx.$$

Thus, the equation

$$\mathrm{P}_{\theta}^{M^{s,n}}(\,dx) = \sqrt{\frac{2}{\pi(2\overline{N}+1)n}} e^{-\frac{2(x-\theta)^2}{(2\overline{N}+1)n}}\,dx$$

holds. Therefore, the sequence of estimators $\vec{M}^s = \{M^{s,n}\}_{n=1}^\infty$ attains the bound $\frac{J_\theta}{2}$ and satisfies the strongly consistent condition. ■

**Proposition 8** *In the half-line quantum Gaussian state family, the sequence of estimators $\vec{M}^w = \{M^{w,n}\}_{n=0}^\infty$ (defined in the following) satisfies the weakly consistent condition and the strongly consistent condition at $\mathbb{R}^+ \setminus \{0\}$ and the relations*

$$\underline{\alpha}(\vec{M}^w, 0) = \alpha(\vec{M}^w, 0) = \frac{\tilde{J}_0}{2} = \log\left(1 + \frac{1}{\overline{N}}\right), \tag{48}$$

$$\underline{\alpha}(\vec{M}^w, \theta) = \alpha(\vec{M}^w, \theta) = \frac{J_\theta}{2} = \frac{1}{\overline{N} + \frac{1}{2}}, \quad \forall \theta \in \mathbb{R}^+ \setminus \{0\}. \tag{49}$$

This proposition indicates the significance of the uniformity of the convergence of (36). This proposition is proven in Appendix D.

[**Construction of** $\vec{M}^w$]: We perform the following unitary evolution:

$$\rho_\theta^{\otimes n-\sqrt{n}} \mapsto \rho_{\sqrt{n}\theta} \otimes \rho_0^{\otimes n-1}.$$

We perform the number measurement $E(N)$ of the first system whose state is $\rho_{\sqrt{n}\theta}$, and let $k$ be its data, where the number operator $N$ is defined as $N := \sum_n n|n\rangle\langle n|$. The estimated value $T_n$ is determined by $T_n := \sqrt{\frac{k}{n}}$. ■

**Theorem 9** *In the one-parameter quantum Gaussian state family, the sequence of estimators* $\vec{M}^w_{\theta_1} = \{M^{w,n}_{\theta_1}\}_{n=1}^\infty$ *(defined in the following) satisfies the weakly consistent condition and the relations*

$$\underline{\alpha}(\vec{M}^w, \theta_1) = \alpha(\vec{M}^w, \theta_1) = \frac{\tilde{J}_\theta}{2} = \log\left(1 + \frac{1}{\overline{N}}\right). \tag{50}$$

[**Construction of** $\vec{M}^w_{\theta_1}$]: We divide $n$ systems into two groups. One consists of $\sqrt{n}$ systems and the other, of $n - \sqrt{n}$ systems. We perform PVM $E(Q)$ for every system in the first group. Let $\xi$ be the mean value in the first group. We perform the following unitary evolution to the second group.

$$\rho_\theta^{\otimes n-\sqrt{n}} \mapsto \rho_{\theta-\theta_1}^{\otimes n-\sqrt{n}}$$

We perform the POVM $M^{w,n-\sqrt{n}}$ to the system whose state is $\rho_{\theta-\theta_1}^{\otimes n-\sqrt{n}}$. Then, we decide the estimated value $\hat{\theta}$ as:

$$\hat{\theta} := \theta_1 + \mathrm{sgn}(\xi - \theta_1) T_{n-\sqrt{n}},$$

where $T_n$ is the estimated value of $M^{w,n}$. ■

*Proof:* Since $\lim_{n\to\infty} \frac{\sqrt{n}}{n} = 0$, similarly to (48), we can prove (50). Also, the weak consistency follows from Theorem 7 and Proposition 8. ■

### *7.2. Finite dimensional family*

In this subsection, we treat the case where the dimension of the Hilbert space $\mathcal{H}$ is $k$ (finite). As for the achievability of inequality (44), we have the following lemma.

**Lemma 10** *Let* $\theta_0$ *be fixed in* $\Theta$*. Assume assumptions 1 and 2. Then, the sequence of estimators* $\vec{M}^s_{\theta_0}$ *(defined in the following) satisfies the strongly consistent condition at* $\theta_0$ *(defined in Def. 5) and the relation*

$$\alpha(\vec{M}^s_{\theta_0}, \theta_0) = \underline{\alpha}(\vec{M}^s_{\theta_0}, \theta_0) = \frac{J_{\theta_0}}{2}. \tag{51}$$

[**Assumption 1**]: The map $\theta \mapsto \rho_\theta$ is $C^1$ and $\rho_\theta > 0$.

[**Assumption 2**]: The map $\theta \mapsto \mathrm{Tr}\, \rho_\theta \frac{L_{\theta_0}}{J_{\theta_0}}$ is injective i.e., one-to-one.

[**Construction of** $\vec{M}^s_{\theta_0}$]: We perform the POVM $E(\frac{L_{\theta_0}}{J_{\theta_0}})$ to all systems. The estimated value is determined to be the mean value plus $\theta_0$. ■

*Proof of Lemma 10:* From assumption 2, the weak consistency is satisfied. Let $\delta > 0$ be a sufficiently small number. Define the function

$$\phi_{\theta,\theta_0}(s) := \operatorname{Tr} \rho_\theta \exp\left( s \left( \frac{L_{\theta_0}}{J_{\theta_0}} - \frac{\operatorname{Tr} \rho_\theta L_{\theta_0}}{J_{\theta_0}} \right) \right). \tag{52}$$

Since $\left\| \frac{L_{\theta_0}}{J_{\theta_0}} \right\| < \infty$ and $\operatorname{Tr} \rho_\theta \left( \frac{L_{\theta_0}}{J_{\theta_0}} - \frac{\operatorname{Tr} \rho_\theta L_{\theta_0}}{J_{\theta_0}} \right) = 0$, we have

$$\lim_{s\to 0} \frac{\phi_{\theta,\theta_0}(s) - 1}{s^2} = \frac{1}{2} \operatorname{Tr} \rho_\theta \left( \frac{L_{\theta_0}}{J_{\theta_0}} - \frac{\operatorname{Tr} \rho_\theta L_{\theta_0}}{J_{\theta_0}} \right)^2.$$

When $\|\theta-\theta_0\|$ is sufficiently small, the function $x \to \sup_s(xs - \log \phi_{\theta,\theta_0}(s))$ is continuous in $(-\delta, \delta)$. Using Cramér's theorem [29], we have

$$\lim_{n\to\infty} \frac{-1}{n} \log \mathrm{P}_\theta^{M_{\theta_0}^{s,n}} \left\{ |\hat\theta - \theta_0| > \epsilon \right\} = \min \left\{ \sup_s(\epsilon s - \log \phi_{\theta,\theta_0}(s)), \sup_{s'}(-\epsilon s' - \log \phi_{\theta,\theta_0}(s')) \right\},$$

for $\epsilon < \delta$. Taking the limit $\epsilon \to 0$, we have

$$\begin{aligned}
&\lim_{\epsilon\to 0}\lim_{n\to\infty} \frac{-1}{\epsilon^2 n} \mathrm{P}_{\theta_0}^{M_{\theta_0}^{s,n}} \{|\hat\theta - \theta_0| > \epsilon\} \\
&= \min \left\{ \lim_{\epsilon\to 0} \frac{\sup_s(\epsilon s - \log \phi_{\theta,\theta_0}(s))}{\epsilon^2}, \lim_{\epsilon\to 0} \frac{\sup_s(-\epsilon s - \log \phi_{\theta,\theta_0}(s))}{\epsilon^2} \right\} = \frac{1}{2} c_{\theta,\theta_0}^{-1}
\end{aligned}$$

where

$$c_{\theta,\theta_0} := \operatorname{Tr} \rho_\theta \left( \frac{L_{\theta_0}}{J_{\theta_0}} - \frac{\operatorname{Tr} \rho_\theta L_{\theta_0}}{J_{\theta_0}} \right)^2$$

because

$$\epsilon s - \log \phi_{\theta,\theta_0}(s) \cong \epsilon s - \log(1 + \frac{1}{2} c_{\theta,\theta_0} s^2) \cong \epsilon s - \frac{1}{2} c_{\theta,\theta_0} s^2 = -\frac{c_{\theta,\theta_0}}{2} \left( s - \frac{\epsilon}{c_{\theta,\theta_0}} \right) + \frac{\epsilon^2}{2 c_{\theta,\theta_0}}.$$

The above convergence is uniform for a neighborhood of $\theta_0$. Taking the limit $\theta \to \theta_0$, we have

$$\lim_{\theta\to\theta_0} \operatorname{Tr} \rho_\theta \left( \frac{L_{\theta_0}}{J_{\theta_0}} - \frac{\operatorname{Tr} \rho_\theta L_{\theta_0}}{J_{\theta_0}} \right)^2 = J_{\theta_0}^{-1} = \operatorname{Tr} \rho_{\theta_0} \left( \frac{L_{\theta_0}}{J_{\theta_0}} - \frac{\operatorname{Tr} \rho_{\theta_0} L_{\theta_0}}{J_{\theta_0}} \right)^2.$$

Thus, we can check (51) and the strong consistency in the neighborhood of $\theta_0$. ∎

But, this sequence of estimators $\vec{M}_\delta^s$ depends on the true parameter $\theta_0$. We should construct a sequence of estimators that satisfies the strongly consistent condition and attains the bound $\frac{J_{\theta_0}}{2}$ at all points $\theta_0$. Since such a construction is too difficult, we introduce another strongly consistent condition that is weaker than the above, and under which inequality (44) holds. We construct a sequence of estimators that satisfies this strongly consistent condition and attains the bound given in (44) for all $\theta$ in a weak sense.

[**Second strongly consistent condition**]: A sequence of estimators $\vec{M} = \{M^n\}$ is called second strongly consistent if there exists a sequence of functions $\{\underline{\beta}_m(\vec{M}, \theta, \epsilon)\}_{m=1}^\infty$ such that

- $\lim_{m\to\infty} \lim_{\epsilon\to 0} \frac{1}{\epsilon^2} \underline{\beta}_m(\vec{M}, \theta, \epsilon) = \underline{\alpha}(\vec{M}, \theta)$.

- $\lim_{\epsilon \to 0} \frac{1}{\epsilon^2} \underline{\beta}_m(\vec{M}, \theta, \epsilon) \le \underline{\alpha}(\vec{M}, \theta)$ holds. Its LHS converges locally uniformly to $\theta$.
- $\forall m, \exists \delta > 0$ s.t. $\underline{\beta}(\vec{M}, \theta, \epsilon) \ge \underline{\beta}_m(\vec{M}, \theta, \epsilon)$, for $\delta > \forall \epsilon > 0$.

Similarly to Theorem 2, we can prove inequality (44) under the second strongly consistent condition.

Under these preparations, we state a theorem with respect to the attainability of the bound $J_\theta$. The following theorem can be regarded as a special case of Theorem 8 of [30].

**Theorem 11** *We assume assumptions 1 and 3. Then, the sequence of estimators* $\vec{M}^s_\delta = \{M^{s,n}_\delta\}_{n=1}^\infty$ *(defined in the following) satisfies the second strongly consistent condition and the relations*

$$\alpha(\vec{M}^s_\delta, \theta) = \underline{\alpha}(\vec{M}^s_\delta, \theta) = (1-\delta)\frac{J_\theta}{2}. \tag{53}$$

*The sequence of estimators* $\vec{M}^s_\delta$ *is independent of the unknown parameter* $\theta$*. Every* $M^{s,n}_\delta$ *is an adaptive estimator and will be defined in section 8.*

Its proof is given in Appendix E.

[**Assumption 3**]: The following set is compact.

$$\left\{ \left( \operatorname{Tr} \rho_\theta \left( \frac{L_{\check{\theta}}}{J_{\check{\theta}}} - \frac{\operatorname{Tr} \rho_\theta L_{\check{\theta}}}{J_{\check{\theta}}} \right)^2 \right)^{-1}, \operatorname{Tr} \rho_\theta \left( \frac{L_{\check{\theta}}}{J_{\check{\theta}}} - \frac{\operatorname{Tr} \rho_\theta L_{\check{\theta}}}{J_{\check{\theta}}} \right)^2 \middle| \forall \theta, \check{\theta} \in \Theta \right\}.$$

If the state family is included by a bounded closed set consisting of positive definite operators, the assumption 3 is satisfied.

[**Construction of** $\vec{M}^s_\delta$]: We perform a faithful POVM $M_f$ (defined in the following) to the first $\delta n$ systems. Then, the data $(\omega_1, \ldots, \omega_{\delta n})$ obey the probability family $\{\mathrm{P}^{M_f}_\theta | \theta \in \Theta\}$. We denote the maximum likelihood estimator (MLE) w.r.t. the data $(\omega_1, \ldots, \omega_{\delta n})$ by $\check{\theta}$. Next, we perform the measurement $E(L_{\check{\theta}})$ defined by the spectral measure of $L_{\check{\theta}}$ to other $(1-\delta)n$ systems. Then, we have data $(\omega_{\delta n+1}, \ldots, \omega_n)$. We decide the final estimated value $T^n_{\check{\theta}}$ as

$$\operatorname{Tr} \rho_{T^n_{\check{\theta}}} L_{\check{\theta}} = \frac{1}{(1-\delta)n} \sum_{i=\delta n+1}^{n} \omega_i.$$

■

**Definition 12** A POVM $M$ is called *faithful*, if the map $\rho \in \mathcal{S}(\mathcal{H}) \mapsto \mathrm{P}^M_\rho$ is one-to-one.

For example, the homodyne measurement $M_h$ is faithful. An example of faithful POVM, which is a POVM taking values in the set of pure states on $\mathcal{H}$, is given by $M_h(d\rho) := k\rho\nu(d\rho)$ , where $\nu$ is the invariant (w.r.t. the action of $\mathrm{SU}(\mathcal{H})$) probability measure on the set of pure states on $\mathcal{H}$. For another example, if $L_1, \ldots L_{k^2-1}$ is a basis of the space of self-adjoint trace-less operators, a disjoint random combination of PVMs $E(L_1), \ldots E(L_{k^2-1})$ is faithful. Note that a disjoint random combination is defined in section 4.

We need to use quantum correlations to achieve the bound $\frac{\tilde{J}_\theta}{2}$. The following theorem can be easily extended to the multi-parameter case.

**Theorem 13** *We assume assumption 1 and that $D(\rho_{\theta'}\|\rho_{\theta_1}) < \infty$ for $\forall \theta' \in \Theta$. Then, the sequence of estimators $\vec{M}^w_{\theta_1} = \{M^{w,n}_{\theta_1}\}_{n=1}^\infty$ satisfies the weakly consistent condition (30), and the equations*

$$\underline{\beta}(\vec{M}^w_{\theta_1}, \theta_1, \epsilon) = \beta(\vec{M}^w_{\theta_1}, \theta_1, \epsilon) = \inf\{D(\rho_{\theta'}\|\rho_{\theta_1}) | |\theta_1 - \theta'| > \epsilon\}, \tag{54}$$

$$\underline{\alpha}(\vec{M}^w_{\theta_1}, \theta_1) = \alpha(\vec{M}^w_{\theta_1}, \theta_1) = \frac{\tilde{J}_{\theta_1}}{2}. \tag{55}$$

*The sequence of estimators $\vec{M}^w_{\theta_1}$ depends on the unknown parameter $\theta_1$ but not on $\epsilon > 0$.*

Its proof is given in Appendix F. In the following construction, $M^{w,n}_{\theta_1}$ is constructed from the PVM $E^n_{\theta_1}$, which is defined from group theoretical viewpoint in Definition 29 in Appendix G.3.

[**Construction of** $M^{w,n}_{\theta_1}$]: We divide the $n$ systems into two groups. We perform a faithful POVM $M_f$ to the first group of $\sqrt{n}$ systems. Then, the data $(\omega_1, \ldots, \omega_{\sqrt{n}})$ obey the probability $\mathrm{P}^{M_f}_\theta$. We let $\check{\theta}$ be the MLE of the data $(\omega_1, \ldots, \omega_{\sqrt{n}})$ under the probability family $\{\mathrm{P}^{M_f}_\theta | \theta \in \Theta\}$. Next, we perform the correlational PVM $E^{n-\sqrt{n}}_{\theta_1}$ to the composite system which consists of the other group of $n - \sqrt{n}$ systems. Then, the data $\omega$ obeys the probability $\mathrm{P}^{E^{n-\sqrt{n}}_{\theta_1}}_\theta$. If $e^{n(1-\delta_{n-\sqrt{n}})D(\rho_{\check{\theta}}\|\rho_{\theta_1})}\mathrm{P}^{E^{n-\sqrt{n}}_{\theta_1}}_{\theta_1}(\omega) \geq \mathrm{P}^{E^{n-\sqrt{n}}_{\theta_1}}_{\check{\theta}}(\omega)$, the estimated value $T_n$ is $\theta_1$, where $\delta_n := \frac{1}{n^{\frac{1}{5}}}$. If not, $T_n$ is $\check{\theta}$. ■

The following lemma proven in Appendix G plays an important role in a proof of Theorem 13.

**Lemma 14** *For three parameters $\theta_0, \theta_1, \theta_2$ and $\delta > 0$, the inequalities*

$$\mathrm{P}^{E^n_{\theta_1}}_{\theta_0}\left\{-\frac{1}{n}\log \mathrm{P}^{E^n_{\theta_1}}_{\theta_2}(\omega) + \mathrm{Tr}\,\rho_{\theta_0}\log\rho_{\theta_2} \geq \delta\right\}$$

$$\leq \exp -n\left(\sup_{0\leq t\leq 1}(\delta - \mathrm{Tr}\,\rho_{\theta_0}\log\rho_{\theta_2})t - t\frac{(k+1)\log(n+1)}{n} - \log \mathrm{Tr}\,\rho_{\theta_0}{\rho_{\theta_2}}^{-t}\right) \tag{56}$$

$$\mathrm{P}^{E^n_{\theta_1}}_{\theta_0}\left\{\frac{1}{n}\log \mathrm{P}^{E^n_{\theta_1}}_{\theta_1}(\omega) - \mathrm{Tr}\,\rho_{\theta_0}\log\rho_{\theta_1} \geq \delta\right\}$$

$$\leq \exp -n\left(\sup_{0\leq t}(\delta + \mathrm{Tr}\,\rho_{\theta_0}\log\rho_{\theta_1})t - \log \mathrm{Tr}\,\rho_{\theta_0}\rho^t_{\theta_1}\right) \tag{57}$$

*hold.*

We obtain the following theorem as a summary of the above discussion.

**Theorem 15** *From Theorem 2, 6, 11 and Lemma 10, we have the equations*

$$\sup_{\vec{M}:\ WC}\limsup_{\epsilon\to 0}\frac{1}{\epsilon^2}\underline{\beta}(\vec{M},\theta,\epsilon) = \sup_{\vec{M}:\ WC}\liminf_{\epsilon\to 0}\frac{1}{\epsilon^2}\underline{\beta}(\vec{M},\theta,\epsilon) = \frac{\tilde{J}_\theta}{2} \tag{58}$$

$$\sup_{\vec{M}:\ SC\ at\ \theta}\liminf_{\epsilon\to 0}\frac{1}{\epsilon^2}\underline{\beta}(\vec{M},\theta,\epsilon) = \frac{J_\theta}{2} \tag{59}$$

*as an operational comparison of $\tilde{J}_\theta$ and $J_\theta$ under assumptions 1, 2 and 3. We can replace $\beta(\vec{M},\theta,\epsilon)$ with $\underline{\beta}(\vec{M},\theta,\epsilon)$ in equations (58).*

As another proof of (29), we can prove (29) from equations (58) and (59).

## 8. Adaptive estimators

In this section, we assume that the dimension of the Hilbert space $\mathcal{H}$ is finite. We consider estimators whose POVM is adaptively chosen from the data. We choose the $l$-th POVM $M_l(\vec{\omega}_{l-1})$ on $\mathcal{H}$ from $l-1$ data $\vec{\omega}_{l-1} := (\omega_1, \ldots, \omega_{l-1})$. Its POVM $M^n$ is described by

$$M^n(\vec{\omega}_n) := M_1(\omega_1) \otimes M_2(\vec{\omega_1}; \omega_2) \otimes \cdots \otimes M_n(\vec{\omega}_{n-1}; \omega_n). \tag{60}$$

In this setting, the estimator is written by the pair $\mathcal{E}_n = (M^n, T_n)$ of the POVM $M^n$ satisfying (60) and the function $T_n : \Omega^n \mapsto \Theta$. Such an estimator $\mathcal{E}_n$ is called an adaptive estimator. As a larger class of POVMs, the separable POVM is well known. A POVM $M^n$ on $\mathcal{H}^{\otimes n}$ is called separable if it is written by

$$M^n = \{M_1(\omega) \otimes \cdots \otimes M_n(\omega)\}_{\omega \in \Omega}$$

on $\mathcal{H}^{\otimes n}$, where $M_i(\omega)$ is a positive semi-definite operator on $\mathcal{H}$. For any separable estimator $(M^n, T_n)$, the relations

$$\begin{aligned}
D^{M^n}(\theta\|\theta') &= \sum_{\omega\in\Omega} \prod_{l'=1}^{n} \operatorname{Tr} \rho_\theta M_{l'}(\omega) \log \frac{\prod_{l=1}^{n} \operatorname{Tr} \rho_\theta M_l(\omega)}{\prod_{l=1}^{n} \operatorname{Tr} \rho_{\theta'} M_l(\omega)} \\
&= \sum_{\omega\in\Omega} \prod_{l'=1}^{n} \operatorname{Tr} \rho_\theta M_{l'}(\omega) \sum_{l=1}^{n} \log \frac{\operatorname{Tr} \rho_\theta M_l(\omega)}{\operatorname{Tr} \rho_{\theta'} M_l(\omega)} \\
&= \sum_{l=1}^{n} \sum_{\omega\in\Omega} a_{\theta,l}(\omega) \operatorname{Tr} \rho_\theta M_l(\omega) \log \frac{a_{\theta,l}(\omega) \operatorname{Tr} \rho_\theta M_l(\omega)}{a_{\theta,l}(\omega) \operatorname{Tr} \rho_{\theta'} M_l(\omega)} \\
&= \sum_{l=1}^{n} D^{M_{\theta,l}}(\theta\|\theta') \le n \sup_{M:\text{POVM on } \mathcal{H}} D^M(\theta\|\theta')
\end{aligned} \tag{61}$$

hold, where the POVM $M_{\theta,l}$ on $\mathcal{H}$ is defined by

$$M_{\theta,l}(\omega) := a_{\theta,l}(\omega) M_l(\omega), \quad a_{\theta,l}(\omega) := \left( \prod_{l'\neq l} \operatorname{Tr} \rho_\theta M_{l'}(\omega) \right).$$

**Theorem 16** *If a sequence of separable estimators $\vec{M} = \{\mathcal{E}_n\} = \{(M^n, T_n)\}$ satisfies the weakly consistent condition, the inequalities*

$$\beta(\vec{M}, \theta_1, \epsilon) \le \inf_{|\theta-\theta_1|>\epsilon} \sup_{M:\text{POVM on } \mathcal{H}} D^M(\theta\|\theta_1) \tag{62}$$

$$\alpha(\vec{M}, \theta_1) \le \frac{J_{\theta_1}}{2} \tag{63}$$

*hold.*

*Proof:* Similarly to (34), the monotonicity of quantum relative entropy yields that

$$-\frac{\log \mathrm{P}_{\theta_1}^{M^n}\{|T_n(\vec{\omega}_n) - \theta_1| > \epsilon\}}{n} \le \frac{D^{M^n}(\theta\|\theta_1) + h(P_n)}{nP_n},$$

where $P_n := \mathrm{P}_\theta^{M^n}\{|T_n(\vec{\omega}_n) - \theta_1| > \epsilon\}$. From the weak consistency, we have $P_n \to 1$. Thus, we obtain (62) from (61). Since $\mathcal{H}$ is finite-dimensional, the set of extremal points of POVMs is compact. Therefore, the convergence $\lim_{\epsilon\to 0} \frac{1}{\epsilon^2} D^M(\theta_1+\epsilon\|\theta_1)$ is uniform w.r.t. $M$. It implies that

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\sup_{M:\text{POVM on }\mathcal{H}} D^M(\theta_1+\epsilon\|\theta_1) = \sup_{M:\text{POVM on }\mathcal{H}} \lim_{\epsilon\to 0}\frac{1}{\epsilon^2} D^M(\theta_1+\epsilon\|\theta_1) = \frac{J_{\theta_1}}{2}. \quad (64)$$

The last equation is derived from (28). ■

The preceding theorem holds for any adaptive estimator. As a simple extension, we can define an $m$-adaptive estimator, that satisfies (60) when every $M_l(\vec{\omega}_{l-1})$ is a POVM on $\mathcal{H}^m$. As a corollary of Theorem 16, we have the following.

**Corollary 17** *If a sequence of $m$-adaptive estimators $\vec{M} = \{\mathcal{E}_n\} = \{(M^n, T_n)\}$ satisfies the weakly consistent condition, then the inequalities*

$$\beta(\vec{M},\theta_1,\epsilon) \le \inf_{|\theta-\theta_1|>\epsilon}\ \sup_{M:\text{POVM on }\mathcal{H}^{\otimes m}} \frac{1}{m} D^M(\theta\|\theta_1) \quad (65)$$

$$\alpha(\vec{M},\theta_1) \quad \le \frac{J_{\theta_1}}{2} \quad (66)$$

*hold.*

Now, we obtain the equation

$$\lim_{m\to\infty}\lim_{\epsilon\to 0}\sup_{\vec{M}:m\text{-AWC}} \frac{1}{\epsilon^2}\underline{\beta}\vec{M},\theta,\epsilon) = \frac{J_\theta}{2}. \quad (67)$$

The part of $\ge$ holds because an adaptive estimator attaining the bound is constructed in Theorem 11, and the part of $\le$ follows from (65) and the equation

$$\begin{aligned}&\lim_{\epsilon\to 0}\sup_{M:\text{POVM on }\mathcal{H}^{\otimes m}} \frac{1}{\epsilon^2 m} D^M(\theta_1+\epsilon\|\theta_1) \\ =& \sup_{M:\text{POVM on }\mathcal{H}^{\otimes m}} \lim_{\epsilon\to 0}\frac{1}{\epsilon^2 m} D^M(\theta_1+\epsilon\|\theta_1) = \frac{J_{\theta_1}}{2},\end{aligned}$$

which is proven similarly to (64).

## 9. Difference in order among limits and supremums

Another operational comparison from theorem 15 is

$$\sup_{\vec{M}:\text{ SC at }\theta}\ \liminf_{\epsilon\to 0}\frac{1}{\epsilon^2}\underline{\beta}(\vec{M},\theta,\epsilon) = \frac{J_\theta}{2} \quad (68)$$

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\sup_{\vec{M}:\text{ SC at }\theta}\underline{\beta}(\vec{M},\theta,\epsilon) \quad = \frac{\tilde{J}_\theta}{2}. \quad (69)$$

Equation (68) equals (59) and equation (69) follows from the following theorem. Therefore, the difference between $\frac{J_\theta}{2}$ and $\frac{\tilde{J}_\theta}{2}$ can be regarded as the difference of the order of $\liminf_{\epsilon\to 0}$ and $\sup_{\vec{M}:\text{ SC}}$.

**Theorem 18** *We assume the assumption 1 in Theorem 11 and that $D(\rho_{\theta'}\|\rho_{\theta_1}) < \infty$ for $\forall \theta' \in \Theta$. For any $\delta > 0$, there exists a sequence $\vec{M}^{m,\delta}_{\theta_0} = \{M^{m,\delta,n}_{\theta_0}\}$ of $m$-adaptive estimators satisfying the strongly consistent condition and the inequality*

$$\begin{aligned}&\lim_{n\to\infty} \frac{-1}{nm} \log \mathrm{P}^{M^{m,\delta,n}_{\theta_0}}_{\theta_0} \{|\hat{\theta}-\theta_0| > \epsilon\} \\ &\ge (1-\delta) \inf \{ D(\theta\|\theta_0) | |\theta-\theta_0| > \epsilon\} - \frac{(1-\delta)(k-1)\log(m+1)}{m}.\end{aligned}$$

However, using Theorem 18, we obtain a stronger equation than (69):

$$\lim_{\epsilon\to 0} \lim_{m\to\infty} \sup_{\vec{M}:m\text{-ASC at }\theta} \frac{1}{\epsilon^2}\beta(\vec{M},\theta,\epsilon) = \frac{\tilde{J}_\theta}{2}, \tag{70}$$

where $m$-ASC at $\theta$ denotes $m$-adaptive and is strongly consistent at $\theta$. This equation is in contrast with (67). Of course, the part of $\le$ for (70) follows from (65). The part of $\ge$ for (70) is derived from the above theorem.

The following two lemmas are essential for our proof of Theorem 18.

**Lemma 19** *For two parameters $\theta_1, \theta_0$, the inequality*

$$mD(\theta_0\|\theta_1) - (k-1)\log(m+1) \le D^{E^m_{\theta_1}}(\theta_0\|\theta_1) \le mD(\theta_0\|\theta_1) \tag{71}$$

*holds, where the PVM $E^m_{\theta_1}$ on $\mathcal{H}^{\otimes m}$ is defined in Appendix G.3. It is independent of $\theta_0$.*

This lemma was proven by Hayashi [21] and can be regarded as an improvement of Hiai and Petz's result [6]. However, Hiai and Petz's original version is sufficient for our proof of Theorem 18. For the reader's convenience, the proof is presented in Appendix G.3.

**Lemma 20** *Let $Y$ be a curved exponential family and $X$ be an exponential family including $Y$. For a curved exponential family and an exponential family, see Chap 4 in Amari and Nagaoka [18] or Barndorff-Nielsen [31]. In this setting, for $n$-i.i.d. data, the MLE $T^{ML}_{X,n}(\omega^n)$ for the exponential family $X$ is sufficient statistic for the curved exponential family $Y$, where $\vec{\omega}_n := (\omega_1,\ldots,\omega_n)$. Using a map $T : X \to Y$, we can define an estimator $T \circ T^{ML}_{X,n}$, and for an estimator $T_Y$, there exists a map $T : X \to Y$ such that $T_Y = T \circ T^{ML}_{X,n}$. We can identify a map $T$ from $X$ to $Y$ with a sequence of estimators $T \circ T^{ML}_{X,n}(\vec{\omega}_n)$. We define the map $T_{\theta_0} : X \to Y$ as:*

$$T_{\theta_0} := \arg\min_{\theta\in Y} \{D(x\|\theta) | D(\theta\|\theta_0) \le D(x\|\theta_0)\}. \tag{72}$$

*When $Y$ is an exponential family (i.e., flat), $T_{\theta_0}$ coincides with the projection to $Y$. Then, the sequence of estimators corresponding to the map $T_{\theta_0}$ satisfies the strong consistency at $\theta_0$ and the equation*

$$\lim_{n\to\infty} \frac{-1}{n} \log p^n_{\theta_0} \{\|T_{\theta_0} \circ T^{ML}_{X,n}(\vec{\omega}_n) - \theta_0\| > \epsilon\} = \inf_{\theta\in Y} \{D(\theta\|\theta_0) | \|\theta-\theta_0\| > \epsilon\}. \tag{73}$$

*holds*

*Proof:* It is well known that for any subset $X' \subset X$, the equation

$$\lim_{n\to\infty} -\frac{1}{n} \log p^n_{\theta_0} \{T^{ML}_{X,n}(\vec{\omega}_n) \in X'\} = \inf_{x\in X'} D(x\|\theta_0) \tag{74}$$

holds. For the reader's convenience, we present a proof of (74) in Appendix H. Thus, equation (73) follows from (72) and (74). If $Y$ is an exponential family, then the estimator $T_{\theta_0} \circ T^{ML}_{X,n}$ coincides with the MLE and satisfies the strong consistency. Otherwise, we choose a neighborhood $U$ of $\theta_0$ so that we can approximate the neighborhood $U$ by the tangent space. The estimator $T_{\theta_0} \circ T^{ML}_{X,n}$ can be approximated by the MLE and satisfies the strong consistency at $U$. Thus, it satisfies the strong consistency at $\theta_0$. ■

*Proof of Theorem 18:* Let $M = \{M_i\}$ be a faithful POVM defined in section 7.2 such that the number of operators $M_i$ is finite. For any $m$ and any $\delta > 0$, we define the disjoint random combination $M^m_{\theta_0}$ of $M \times m$ and $E^m_{\theta_0}$ in the ratio $\delta : 1 - \delta$. Note that a disjoint random combination is defined in section 4. From the definition of $M^m_{\theta_0}$, the inequality

$$(1-\delta) D^{E^m_{\theta_0}}(\theta\|\theta) \le D^{M^m_{\theta_0}}(\theta\|\theta) \tag{75}$$

holds. Since the map $\theta \mapsto \mathrm{P}^M_\theta$ is one-to-one, the map $\theta \mapsto \mathrm{P}^{M^m_{\theta_0}}_\theta$ is also one-to-one. Since $M$ and $E^m_{\theta_0}$ are finite-resolutions of the identity, the one-parameter family $\{\mathrm{P}^{M^m_{\theta_0}}_\theta | \theta \in \Theta\}$ is a subset of multi-nominal distributions $X$, which is an exponential family. Applying Lemma 20, we have

$$\begin{aligned}
& \lim_{n\to\infty} \frac{-1}{nm} \log \mathrm{P}^{M^m_{\theta_0}\times n}_{\theta_0} \{ |T_{\theta_0} \circ T^{ML}_{X,n}(\vec{\omega}_n) - \theta_0| > \epsilon \} \\
= & \frac{1}{M} \inf_{\theta\in\Theta} \{ D^{M^m_{\theta_0}}(\theta\|\theta_0) \| |\theta - \theta_0| > \epsilon \} \\
\ge & \frac{(1-\delta)}{m} \inf \left\{ D^{E^m_{\theta_0}}(\theta\|\theta_0) \middle| |\theta - \theta_0| > \epsilon \right\} \\
\ge & (1-\delta) \inf \{ D(\theta\|\theta_0) | |\theta - \theta_0| > \epsilon \} - \frac{(1-\delta)(k-1)\log(m+1)}{m},
\end{aligned}$$

where the first inequality follows from (75) and the second inequality follows from (71). ■

**Remark 2** In the case of one-parameter equatorial spin 1/2 system state family, the map $\theta \mapsto \mathrm{P}^{E^m_{\theta_0}}_\theta$ is not one-to-one. Therefore, we need to treat not $E^m_{\theta_0}$ but $M^m_{\theta_0}$.

## Conclusions

It has been clarified that the SLD Fisher information $J_\theta$ gives the essential large deviation bound in the quantum estimation and the KMB Fisher information $\tilde{J}_\theta$ gives the large deviation bound of superefficient estimators. Since estimators attaining the bound $\frac{\tilde{J}_\theta}{2}$ is unnatural, the bound $\frac{J_\theta}{2}$ is more important from the viewpoint of quantum estimation than the bound $\frac{\tilde{J}_\theta}{2}$. On the other hand, concerning a quantum analogue of information geometry from the viewpoint of e-connections, KMB is most natural among the quantum versions of the Fisher information. The interpretation of these two facts which seem to contradict each other, remains a problem. Similarly, it is a future problem to understand geometrically the relationship between the change of the orders of limits and the difference between the two quantum analogues of the Fisher information.

**Acknowledgments**

The author wishes to thank Professor H. Nagaoka for encouragement and essentially advice regarding this manuscript He also wishes to thank Professor K. Matsumoto for useful advice regarding this manuscript.

## Appendix A. Proof of Lemma 3

We define the unitary operator $U_\epsilon$ as

$$b^2(\rho_\theta, \rho_{\theta+\epsilon}) = 2\,(1 - \mathrm{Tr}\,|\sqrt{\rho_\theta}\sqrt{\rho_{\theta+\epsilon}}|) = \mathrm{Tr}(\sqrt{\rho} - \sqrt{\sigma}U_\epsilon)(\sqrt{\rho} - \sqrt{\sigma}U_\epsilon)^*.$$

Letting $W(\epsilon)$ be $\sqrt{\rho_{\theta+\epsilon}}U_\epsilon$, then we have

$$\begin{aligned} b^2(\rho_\theta, \rho_{\theta+\epsilon}) =& \ \mathrm{Tr}(W(0) - W(\epsilon))(W(0) - W(\epsilon))^* \\ \cong& \ \mathrm{Tr}\left(-\frac{dW}{dt}(0)\epsilon\right)\left(-\frac{dW}{dt}(0)\epsilon\right)^* \cong \mathrm{Tr}\,\frac{dW}{dt}(0)\frac{dW}{dt}(0)^*\epsilon^2. \end{aligned}$$

As is proven in the following discussion, the SLD $L$ satisfies that

$$\frac{dW}{dt}(0) = \frac{1}{2}LW(0). \tag{A.1}$$

Therefore, we have

$$b^2(\rho_\theta, \rho_{\theta+\epsilon}) \cong \mathrm{Tr}\,\frac{1}{4}LW(0)W(0)^*L\epsilon^2 = \frac{1}{4}\,\mathrm{Tr}\,L^2\rho_\theta\epsilon.$$

We obtain (37). It is sufficient to show (A.1).

From the definition of Bures distance, we have

$$\begin{aligned} b^2(\rho_\theta, \rho_{\theta+\epsilon}) &= \min_{U:unitary} \mathrm{Tr}(\sqrt{\rho_\theta} - \sqrt{\rho_{\theta+\epsilon}}U)(\sqrt{\rho_\theta} - \sqrt{\rho_{\theta+\epsilon}}U)^* \\ &= 2 - \max_{U:unitary} \mathrm{Tr}\,\sqrt{\rho_\theta}\sqrt{\rho_{\theta+\epsilon}}U^* + U\sqrt{\rho_{\theta+\epsilon}}\sqrt{\rho_\theta} \\ &= 2 - \mathrm{Tr}\,|\sqrt{\rho_\theta}\sqrt{\rho_{\theta+\epsilon}}| + |\sqrt{\rho_{\theta+\epsilon}}\sqrt{\rho_\theta}| \\ &= 2 - \mathrm{Tr}\left(\sqrt{\rho_\theta}\sqrt{\rho_{\theta+\epsilon}}U(\epsilon)^* + U(\epsilon)\sqrt{\rho_{\theta+\epsilon}}\sqrt{\rho_\theta}\right). \end{aligned}$$

It implies that $\sqrt{\rho_\theta}\sqrt{\rho_{\theta+\epsilon}}U(\epsilon)^* = U(\epsilon)\sqrt{\rho_{\theta+\epsilon}}\sqrt{\rho_\theta}$. Therefore, $W(0)W(\epsilon)^* = W(\epsilon)W(0)^*$. Taking the derivative, we have

$$W(0)\frac{dW}{d\epsilon}(0)^* = \frac{dW}{d\epsilon}(0)W(0)^*.$$

It implies that there exists a self-adjoint operator $L$ such that

$$\frac{dW}{d\epsilon}(0) = \frac{1}{2}LW(0).$$

Since $\rho_{\theta+\epsilon} = W(\epsilon)W(\epsilon)^*$, we have

$$\frac{d\rho}{d\theta}(\theta) = \frac{1}{2}\left(LW(0)W(0)^* + W(0)W(0)^*L\right).$$

Thus, the operator $L$ coincides with the SLD.

## Appendix B. Proof of (42)

Let $M = \{M_i\}$ be an arbitrary POVM. We choose the unitary $U$ satisfying that

$$U\sigma^{1/2}\rho^{1/2} = \sqrt{\rho^{1/2}\sigma\rho^{1/2}}.$$

Using the Schwarz inequality, we have

$$\begin{aligned}\sqrt{\mathrm{P}^M_\rho(\omega)}\sqrt{\mathrm{P}^M_\sigma(\omega)} &= \sqrt{\mathrm{Tr}\left(M_\omega^{1/2}\sigma^{1/2}U^*\right)^*\left(M_\omega^{1/2}\sigma^{1/2}U^*\right)}\sqrt{\mathrm{Tr}\left(M_\omega^{1/2}\rho^{1/2}\right)^*\left(M_\omega^{1/2}\rho^{1/2}\right)} \\ &\geq \mathrm{Tr}\left(M_\omega^{1/2}\sigma^{1/2}U^*\right)^*\left(M_\omega^{1/2}\rho^{1/2}\right) = \left|\mathrm{Tr}\, U\sigma^{1/2}M_\omega\rho^{1/2}\right|.\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_\omega \sqrt{\mathrm{P}^M_\rho(\omega)}\sqrt{\mathrm{P}^M_\sigma(\omega)} &\geq \sum_\omega \left|\mathrm{Tr}\, U\sigma^{1/2}M_\omega\rho^{1/2}\right| \geq \left|\sum_\omega \mathrm{Tr}\, U\sigma^{1/2}M_\omega\rho^{1/2}\right| \\ &= \left|\mathrm{Tr}\, U\sigma^{1/2}\rho^{1/2}\right| = \mathrm{Tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}.\end{aligned}$$

## Appendix C. Proof of Lemma 4

Let $m$ and $\epsilon$ be an arbitrary positive integer and an arbitrary positive real number, respectively. There exists a sufficiently large integer $N$ such that

$$\begin{aligned}&\frac{1}{n}\log \mathrm{P}^{M^n}_\theta\left\{|\hat\theta-\theta| > \frac{\delta}{m}i\right\} \leq -\underline{\beta}\left(\vec{M},\theta,\frac{\delta}{m}i\right)+\epsilon \\ &\frac{1}{n}\log \mathrm{P}^{M^n}_{\theta+\delta}\left\{|\hat\theta-(\theta+\delta)| > \frac{\delta}{m}(m-i)\right\} \leq -\underline{\beta}\left(\vec{M},\theta+\delta,\frac{\delta}{m}(m-i)\right)+\epsilon\end{aligned}$$

for $i = 0,\ldots,m$ and $\forall n \geq N$. From the monotonicity (41) and the additivity (38) of quantum affinity, we can evaluate as follows:

$$\begin{aligned}&-\frac{n}{8}I(\rho_\theta\|\rho_{\theta+\delta}) = -\frac{1}{8}I(\rho_\theta^{\otimes n}\|\rho_{\theta+\delta}^{\otimes n}) \\ \leq\ &\log\Bigg(\mathrm{P}^{M^n}_\theta\left\{\hat\theta\leq\theta\right\}^{\frac12}\mathrm{P}^{M^n}_{\theta+\delta}\left\{\hat\theta\leq\theta\right\}^{\frac12} + \mathrm{P}^{M^n}_\theta\left\{\theta+\delta\ <\hat\theta\right\}^{\frac12}\mathrm{P}^{M^n}_{\theta+\delta}\left\{\theta+\delta\ <\hat\theta\right\}^{\frac12} \\ &+\sum_{i=1}^m \mathrm{P}^{M^n}_\theta\left\{\theta+\frac{\delta}{m}(i-1)\ <\hat\theta\leq\theta+\frac{\delta}{m}i\right\}^{\frac12}\mathrm{P}^{M^n}_{\theta+\delta}\left\{\theta+\frac{\delta}{m}(i-1)\ <\hat\theta\leq\theta+\frac{\delta}{m}i\right\}^{\frac12}\Bigg) \\ \leq\ &\log\Bigg(\mathrm{P}^{M^n}_{\theta+\delta}\left\{\left|\hat\theta-(\theta+\delta)\right|\geq\delta\right\}^{\frac12} + \mathrm{P}^{M^n}_\theta\left\{\left|\hat\theta-\theta\right|\ >\delta\right\}^{\frac12} \\ &+\sum_{i=1}^m \mathrm{P}^{M^n}_\theta\left\{\left|\hat\theta-\theta\right|\ >\frac{\delta}{m}(i-1)\right\}^{\frac12}\mathrm{P}^{M^n}_{\theta+\delta}\left\{\left|\hat\theta-(\theta+\delta)\right|\geq\frac{\delta}{m}(m-i)\right\}^{\frac12}\Bigg) \\ \leq\ &\log\Bigg(\mathrm{P}^{M^n}_{\theta+\delta}\left\{\left|\hat\theta-(\theta+\delta)\right|\ >\frac{\delta}{m}(m-1)\delta\right\}^{\frac12} + \mathrm{P}^{M^n}_\theta\left\{\left|\hat\theta-\theta\right|\ >\delta\right\}^{\frac12} \\ &+\sum_{i=1}^m \mathrm{P}^{M^n}_\theta\left\{\left|\hat\theta-\theta\right|\ >\frac{\delta}{m}(i-1)\right\}^{\frac12}\mathrm{P}^{M^n}_{\theta+\delta}\left\{\left|\hat\theta-(\theta+\delta)\right|\ >\frac{\delta}{m}(m-i-1)\right\}^{\frac12}\Bigg)\end{aligned}$$

$$
\begin{aligned}
&\le \log\Bigg( \exp\left(-\frac{n}{2}\left(\underline{\beta}\left(\vec{M},\theta,\frac{\delta}{m}(m-1)\right)-\epsilon\right)\right) + \exp\left(-\frac{n}{2}\left(\underline{\beta}\left(\vec{M},\theta+\delta,\delta\right)-\epsilon\right)\right) \\
&\quad + \sum_{i=1}^{m} \exp\left(-\frac{n}{2}\left(\underline{\beta}\left(\vec{M},\theta,\frac{\delta}{m}(i-1)\right)-\epsilon\right) - \frac{n}{2}\left(\underline{\beta}\left(\vec{M},\theta+\delta,\frac{\delta}{m}(m-i-1)\right)-\epsilon\right)\right)\Bigg) \\
&\le \log(m+2)\exp\left(-\frac{n}{2}\min_{0\le i\le m}\left(\underline{\beta}\left(\vec{M},\theta,\frac{\delta}{m}(i-1)\right)+\underline{\beta}\left(\vec{M},\theta+\delta,\frac{\delta}{m}(m-i-1)\right)-2\epsilon\right)\right) \\
&= \log(m+2) - \frac{n}{2}\left(\min_{0\le i\le m}\underline{\beta}\left(\vec{M},\theta,\frac{\delta}{m}(i-1)\right)+\underline{\beta}\left(\vec{M},\theta+\delta,\frac{\delta}{m}(m-i-1)\right)-2\epsilon\right),
\end{aligned}
$$

where we assume that $\underline{\beta}(\vec{M},\theta,a)=0$ for any negative real number $a$. Taking the limit $n\to\infty$ after dividing by $n$, we have

$$
\frac{1}{8}I(\rho_\theta\|\rho_{\theta+\delta}) \ge \frac{1}{2}\min_{0\le i\le m}\left(\underline{\beta}\left(\vec{M},\theta,\frac{\delta}{m}(i-1)\right)+\underline{\beta}\left(\vec{M},\theta+\delta,\frac{\delta}{m}(m-i-1)\right)-2\epsilon\right).
$$

Since $\epsilon>0$ is arbitrary, the inequality

$$
\frac{1}{8}I(\rho_\theta\|\rho_{\theta+\delta}) \ge \frac{1}{2}\min_{0\le i\le m}\left(\underline{\beta}\left(\vec{M},\theta,\frac{\delta}{m}(i-1)\right)+\underline{\beta}\left(\vec{M},\theta+\delta,\frac{\delta}{m}(m-i-1)\right)\right)
$$

holds. Taking the limit $m\to\infty$, we obtain (43).

## Appendix D. Proof of Proposition 8

For a proof of Proposition 8, we need the following lemma.

**Lemma 21** *Let $g_n(\omega), f_n(\omega)$ be functions on $\Omega$. Assume that the functions $\beta_1(\omega) := \lim_{n\to\infty}\frac{-1}{n}\log f_n(\omega), \beta_2(\omega) := \lim_{n\to\infty}\frac{-1}{n}\log g_n(\omega)$ are continuous. If the inequality $g_n(\omega)\le 1$ holds for any element $\omega\in\Omega$ and any positive integer $n$, and if there exists a subset $K\subset\Omega$ such that*

$$
\lim_{n\to\infty}\frac{-1}{n}\log\left(\int_K f_n(\omega)\,d\omega\right) > \min_{\omega\in\Omega}\beta_1(\omega)+\beta_2(\omega),
$$

*the relation*

$$
\lim_{n\to\infty}\frac{-1}{n}\log\left(\int_\Omega f_n(\omega)g_n(\omega)\,d\omega\right) = \min_{\omega\in\Omega}\beta_1(\omega)+\beta_2(\omega)
$$

*holds.*

Similarly to Lemma 4, Lemma 21 is proven.

Now, we will prove Proposition 8. Form the definition of $M^{w,n}$ and the equation $\rho_0=\frac{1}{\overline{N}+1}\sum_k\left(\frac{\overline{N}}{\overline{N}+1}\right)^k|k\rangle\langle k|$, we have

$$
\log \mathrm{P}_0^{M^{s,n}}\{T_n>\epsilon\} = \log\sum_{k>n\epsilon^2}\left(\frac{\overline{N}}{\overline{N}+1}\right)^k = \log\left(\frac{\overline{N}}{\overline{N}+1}\right)^{[n\epsilon^2]},
$$

where [ ] is a Gauss notation. Therefore, we obtain

$$
\beta(\vec{M}^w,0,\epsilon)=\epsilon^2\log\left(1+\frac{1}{\overline{N}}\right),
$$

which implies (48).

Next, we prove the strongly consistent condition and (49). We can calculate as follows:

$$\begin{aligned}
\mathrm{P}_\theta^{M^{w,n}}\{T_n - \theta > \epsilon\} &= \sum_{k > (\theta+\epsilon)^2 n} \langle k| \int_{\mathbb{C}} \frac{1}{\pi \overline{N}} |\alpha\rangle\langle\alpha| e^{-\frac{|\alpha - \sqrt{n}\theta|^2}{\overline{N}}} \, d^2\alpha |k\rangle \\
&= \int_{\mathbb{C}} \frac{\sqrt{n}}{\pi\overline{N}} e^{-n\frac{|\alpha-\theta|^2}{\overline{N}}} \sum_{k > (\theta-\epsilon)^2 n} \frac{(n|\alpha|^2)^k}{k!} e^{-n|\alpha|^2} \, d^2\alpha. \quad \text{(D.1)}
\end{aligned}$$

The equation

$$\lim_{n\to\infty} \frac{-1}{n} \log \frac{\sqrt{n}}{\pi\overline{N}} e^{-n\frac{|\alpha-\theta|^2}{\overline{N}}} = \frac{|\alpha-\theta|^2}{\overline{N}} \quad \text{(D.2)}$$

holds. Also, as is proven in the latter, the equations

$$\begin{aligned}
&\lim_{n\to\infty} \frac{-1}{n} \log \left( \sum_{k > (\theta+\epsilon)^2 n} \frac{(n|\alpha|^2)^k}{k!} e^{-n|\alpha|^2} \right) \\
&= \left( (\theta+\epsilon)^2 \log \frac{(\theta+\epsilon)^2}{|\alpha|^2} + |\alpha|^2 - (\theta+\epsilon)^2 \right) 1((\theta+\epsilon)^2 - |\alpha|^2) \quad \text{(D.3)} \\
&\lim_{n\to\infty} \frac{-1}{n} \log \left( \sum_{k < (\theta-\epsilon)^2 n} \frac{(n|\alpha|^2)^k}{k!} e^{-n|\alpha|^2} \right) \\
&= \left( (\theta-\epsilon)^2 \log \frac{(\theta-\epsilon)^2}{|\alpha|^2} + |\alpha|^2 - (\theta-\epsilon)^2 \right) 1(-(\theta-\epsilon)^2 + |\alpha|^2) \quad \text{(D.4)}
\end{aligned}$$

hold, where $1(x)$ is defined as

$$1(x) = \begin{cases} 1 & x \ge 0 \\ 0 & x < 0. \end{cases}$$

For any $\delta > 0$, there exists a real number $K$ such that

$$\lim_{n\to\infty} -\frac{1}{n} \log \left( \int_{|\alpha| > K} \frac{\sqrt{n}}{\pi\overline{N}} \exp\left( -n \frac{|\alpha-\theta|^2}{\overline{N}} \right) dx \right) = \frac{K-\theta}{\overline{N}} > \delta.$$

Now, we can apply Lemma 21 to (D.1). From (D.2) and (D.3), the relations

$$\begin{aligned}
&\lim_{n\to\infty} \frac{-1}{n} \log \mathrm{P}_\theta^{M^{w,n}}\{T_n - \theta > \epsilon\} \\
&= \min_{\alpha\in\mathbb{C}} \frac{|\alpha-\theta|^2}{\overline{N}} + \left( (\theta+\epsilon)^2 \log \frac{(\theta+\epsilon)^2}{|\alpha|^2} + |\alpha|^2 - (\theta+\epsilon)^2 \right) 1((\theta+\epsilon)^2 - |\alpha|^2) \\
&= \min_{\alpha\in\mathbb{R}} \frac{|\alpha-\theta|^2}{\overline{N}} + \left( (\theta+\epsilon)^2 \log \frac{(\theta+\epsilon)^2}{|\alpha|^2} + |\alpha|^2 - (\theta+\epsilon)^2 \right) 1((\theta+\epsilon)^2 - |\alpha|^2) \\
&= \min_{s\in\mathbb{R}} \frac{s^2}{\overline{N}} + \left( (\theta+\epsilon)^2 \log \frac{(\theta+\epsilon)^2}{(\theta-s)^2} + (\theta-s)^2 - (\theta+\epsilon)^2 \right) 1((\theta+\epsilon)^2 - (\theta-s)^2)
\end{aligned}$$

hold. If $\epsilon$ is sufficiently small for $\theta$, we have the following approximation:

$$\lim_{n\to\infty} \frac{-1}{n} \log \mathrm{P}_\theta^{M^{w,n}}\{T_n - \theta > \epsilon\} \cong \min_s \frac{1+2\overline{N}}{\overline{N}} \left( s - \frac{2\overline{N}}{1+2\overline{N}} \epsilon \right)^2 + \frac{\epsilon^2}{\overline{N} + \frac{1}{2}}.$$

Thus,

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{-1}{n\epsilon^2}\log \mathrm{P}_\theta^{M^{w,n}}\{T_n-\theta\ >\epsilon\}=\frac{1}{\overline{N}+\frac{1}{2}}. \tag{D.5}$$

The second convergence of LHS of (D.5) is uniform at a sufficiently small neighborhood $U_{\theta_0}$ of arbitrary $\theta_0\in\mathbb{R}^+\setminus\{0\}$.

Similarly to (D.5), from (D.4), we can prove

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{-1}{n\epsilon^2}\log \mathrm{P}_\theta^{M^{w,n}}\{T_n-\theta\ <-\epsilon\}=\frac{1}{\overline{N}+\frac{1}{2}}. \tag{D.6}$$

Also, the second convergence of LHS of (D.6) is uniform at a sufficiently small neighborhood $U_{\theta_0}$ of arbitrary $\theta_0\in\mathbb{R}^+\setminus\{0\}$. Thus, (49) and the strongly consistent condition are proven.

Next, we prove (D.3) and (D.4). Using the Stirling formula, we have

$$\lim_{n\to\infty}\frac{-1}{n}\log\frac{(n|\alpha|^2)^{[\delta n]}}{[\delta n]!}e^{-n|\alpha|^2}=\left(\delta\log\frac{\delta}{|\alpha|^2}+|\alpha|-\delta^2\right)1(\delta-|\alpha|^2). \tag{D.7}$$

Since the relations

$$\frac{(n|\alpha|^2)^{([(\theta-\epsilon)^2n]-1)}}{([(\theta-\epsilon)^2n]-1)!}e^{-n|\alpha|^2}\le\sum_{k\,<(\theta-\epsilon)^2n}\frac{(n|\alpha|^2)^k}{k!}e^{-n|\alpha|^2}\le[(\theta-\epsilon)^2n]\frac{(n|\alpha|^2)^{([(\theta-\epsilon)^2n]-1)}}{([(\theta-\epsilon)^2n]-1)!}e^{-n|\alpha|^2}$$

hold, (D.4) follows from (D.7). If $(\theta+\epsilon)^2\le|\alpha|^2$, the equation

$$\lim_{n\to\infty}\frac{-1}{n}\log\sum_{k\,>(\theta+\epsilon)^2n}\frac{(n|\alpha|^2)^k}{k!}e^{-n|\alpha|^2}=0 \tag{D.8}$$

holds. It implies (D.3) in the case of $(\theta+\epsilon)^2\le|\alpha|^2$.

Next we prove (D.3) in the case of $(\theta+\epsilon)^2\ >|\alpha|^2$. In this case, we have

$$\sum_{Ln\,>k\,>(\theta+\epsilon)^2n}\frac{(n|\alpha|^2)^k}{k!}e^{-n|\alpha|^2}\le n(L-(\theta+\epsilon)^2)\frac{(n|\alpha|^2)^{[(\theta+\epsilon)^2n]}}{[(\theta+\epsilon)^2n]!}e^{-n|\alpha|^2} \tag{D.9}$$

because $\left(\frac{(n|\alpha|^2)^k}{k!}e^{-n|\alpha|^2}\right)/\left(\frac{(n|\alpha|^2)^{(k+1)}}{(k+1)!}e^{-n|\alpha|^2}\right)=\frac{k+1}{n|\alpha|^2}$. If $L$ and $N$ are sufficiently large for $|\alpha|^2$, we have

$$\sum_{k\ge Ln}\frac{(n|\alpha|^2)^k}{k!}e^{-n|\alpha|^2}\le\sum_{k\ge Ln}e^{-k}=\frac{e^{-nL}}{1-e^{-1}} \tag{D.10}$$

because (D.7) implies that

$$\frac{(n|\alpha|^2)^{[\delta n]}}{[\delta n]!}e^{-n|\alpha|^2}\le e^{-[\delta n]},\quad\forall\delta\ge L,\forall n\ge N.$$

Since the relations

$$\begin{aligned}\frac{(n|\alpha|^2)^{[(\theta+\epsilon)^2n]}}{[(\theta+\epsilon)^2n]!}e^{-n|\alpha|^2}\le&\sum_{k\,>(\theta+\epsilon)^2n}\frac{(n|\alpha|^2)^k}{k!}e^{-n|\alpha|^2}\\&\le n(L-(\theta+\epsilon)^2)\frac{(n|\alpha|^2)^{[(\theta+\epsilon)^2n]}}{[(\theta+\epsilon)^2n]!}e^{-n|\alpha|^2}+\frac{e^{-nL}}{1-e^{-1}}\end{aligned}$$

hold, we have

$$
\begin{aligned}
&\left((\theta+\epsilon)^2\log\frac{(\theta+\epsilon)^2}{|\alpha|^2}+|\alpha|^2-(\theta+\epsilon)^2\right)\\
\geq\ &\lim_{n\to\infty}\frac{-1}{n}\log\left(\sum_{k>(\theta+\epsilon)^2 n}\frac{(n|\alpha|^2)^k}{k!}e^{-n|\alpha|^2}\right)\\
\geq\ &\min\left\{\left((\theta+\epsilon)^2\log\frac{(\theta+\epsilon)^2}{|\alpha|^2}+|\alpha|^2-(\theta+\epsilon)^2\right),L\right\}.
\end{aligned}
$$

If we Let $L$ be a sufficiently large real number, we have (D.3).

## Appendix E. Proof of Theorem 11

In this proof, we use the function $\phi_{\theta,\check{\theta}}(s)$ defined in (H.1). First, we prove the following four facts.

**(i)** The faithful POVM $M_f$ satisfies the inequalities

$$
\beta(\vec{M}_f,\theta,\epsilon)>0,\quad \alpha(\vec{M}_f,\theta)>0.
$$

**(ii)** The relation

$$
\lim_{\check{\theta}\to\theta}\left(\operatorname{Tr}\rho_\theta\left(\frac{L_{\check{\theta}}}{J_{\check{\theta}}}-\frac{\operatorname{Tr}\rho_\theta L_{\check{\theta}}}{J_{\check{\theta}}}\right)^2\right)^{-1}=J_\theta,\quad \forall\theta\in\Theta
$$

holds.

**(iii)** The equation

$$
\lim_{s\to 0}\frac{\phi_{\theta,\check{\theta}}(s)-1}{s^2}=\frac{1}{2}\operatorname{Tr}\rho_\theta\left(\frac{L_{\check{\theta}}}{J_{\check{\theta}}}-\frac{\operatorname{Tr}\rho_\theta L_{\check{\theta}}}{J_{\check{\theta}}}\right)^2 \tag{E.1}
$$

holds. The LHS converges uniformly w.r.t. $\theta,\check{\theta}$.

**(iv)** For any real number $\delta_2>0$, there exists a sufficiently small real number $\epsilon>0$ such that if $|\operatorname{Tr}\rho_\theta L_{\check{\theta}}-\operatorname{Tr}\rho_{\theta'}L_{\check{\theta}}|\leq\epsilon(1-\delta_2)$ and $|\check{\theta}-\theta|<\sqrt{\epsilon}$, then $|\theta'-\theta|<\epsilon$.

Fact (i) is easily proven from the definition of $M_f$. Fact (iii) is proven by the relation

$$
\sup_{\check{\theta},\theta}\left\|\frac{L_{\check{\theta}}}{J_{\check{\theta}}}-\frac{\operatorname{Tr}\rho_\theta L_{\check{\theta}}}{J_{\check{\theta}}}\right\|<\infty.
$$

Fact (ii) is, also, proven by the relations

$$
\operatorname{Tr}\rho_\theta\left(\frac{L_{\check{\theta}}}{J_{\check{\theta}}}-\frac{\operatorname{Tr}\rho_\theta L_{\check{\theta}}}{J_{\check{\theta}}}\right)^2=\frac{\operatorname{Tr}\rho_\theta\left(L_{\check{\theta}}^2\right)}{J_{\check{\theta}}^2}-\frac{(\operatorname{Tr}\rho_\theta L_{\check{\theta}})^2}{J_{\check{\theta}}^2}\to J_\theta^{-1}\text{ as }\check{\theta}\to\theta.
$$

Fact (iv) follows from the relation

$$
\frac{\partial\operatorname{Tr}\rho_\theta L_{\check{\theta}}}{\partial\theta}\to 1\text{ as }\check{\theta}\to\theta,
$$

which follows from fact (i).

Next, we prove the theorem from the preceding four facts. The inequality

$$\begin{aligned}&\mathrm{P}_\theta^{M_\delta^{s,n}}\{\hat\theta \notin U_{\theta,\epsilon}\}\\ \le\ &\mathrm{P}_\theta^{M_f\times\delta n}\{\hat\theta \in U_{\theta,\sqrt{\epsilon}}\} \sup_{\breve\theta\in U_{\theta,\sqrt{\epsilon}}} \mathrm{P}_\theta^{L_{\breve\theta}\times(1-\delta)n}\{\hat\theta\notin U_{\theta,\epsilon}\} + \mathrm{P}_\theta^{M_f\times\delta n}\{\hat\theta \notin U_{\theta,\sqrt{\epsilon}}\}\end{aligned} \tag{E.2}$$

holds. As is proven in the latter, the inequality

$$\begin{aligned}&\liminf_{n\to\infty} -\frac{1}{n}\log \sup_{\breve\theta\in U_{\theta,\sqrt{\epsilon}}} \mathrm{P}_\theta^{L_{\breve\theta}\times(1-\delta)n}\left\{T_{\breve\theta}^n \notin U_{\theta,\epsilon}\right\}\\ \ge\ &(1-\delta)g\left(\epsilon^2(1-\delta_2)^2\frac{1}{2}\left(\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\breve\theta}}{J_{\breve\theta}}-\frac{\mathrm{Tr}\,\rho_\theta L_{\breve\theta}}{J_{\breve\theta}}\right)^2\right)^{-1}, \frac{\epsilon^2(1-\delta_2)^2}{2}\delta\right)\end{aligned} \tag{E.3}$$

holds, where the function $g(x,y)$ is defined as $g(x,y) := x - \log(1+\frac{x}{2}+y)$. Therefore, we have

$$\begin{aligned}&\underline{\beta}(\vec{M}_\delta^s,\theta,\epsilon) = \liminf_{n\to\infty} -\frac{1}{n}\log \mathrm{P}_\theta^{M_\delta^{s,n}}\{\hat\theta\notin U_{\theta,\sqrt{\epsilon}}\}\\ \ge\ &\min\Bigg\{(1-\delta)h\left(\epsilon^2(1-\delta_2)^2\frac{1}{2}\left(\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\breve\theta}}{J_{\breve\theta}}-\frac{\mathrm{Tr}\,\rho_\theta L_{\breve\theta}}{J_{\breve\theta}}\right)^2\right)^{-1}, \frac{\epsilon^2(1-\delta_2)^2}{2}\delta\right),\\ &\quad c\underline{\beta}(\{M_f\times\delta n\},\theta,\sqrt{\epsilon})\Bigg\}.\end{aligned} \tag{E.4}$$

From facts (i) and (ii), the equations

$$\begin{aligned}&\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}(\text{RHS of (E.4)})\\ =\ &\frac{1-\delta}{2}\left(\lim_{\breve\theta\to\theta}(1-\delta_1)^2(1-\delta_2)^2\left(\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\breve\theta}}{J_{\breve\theta}}-\frac{\mathrm{Tr}\,\rho_\theta L_{\breve\theta}}{J_{\breve\theta}}\right)^2\right)^{-1} - (1-\delta_2)^2\delta_3\right)\\ =\ &\frac{1-\delta}{2}\left((1-\delta_1)^2(1-\delta_2)^2 J_\theta - (1-\delta_2)^2\delta_3\right)\end{aligned} \tag{E.5}$$

hold. The RHS of (E.5) converges locally uniformly w.r.t. $\theta$. Let $\underline{\beta}_m(\vec{M}_\delta^s,\theta,\epsilon)$ be the RHS of (E.4) in the case of $\delta_2=\delta_3=\frac{1}{m}$. Therefore, we have

$$\lim_{m\to\infty}\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\underline{\beta}_m(\vec{M}_\delta^s,\theta,\epsilon) = \frac{1-\delta}{2}J_\theta.$$

It implies that

$$\underline{\alpha}(\vec{M}_\delta^s,\theta) \ge \frac{1-\delta}{2}J_\theta.$$

If the converse inequality

$$\alpha(\vec{M}_\delta^s,\theta) \le \frac{1-\delta}{2}J_\theta \tag{E.6}$$

holds, we can immediately show the relations (53) and that the sequence of estimators $\vec{M}_\delta^s$ satisfies the second strongly consistent condition.

In the following, the relations (E.6) and (E.3) are proven. First, we prove (E.6). We can evaluate the probability $\mathrm{P}_\theta^{M_\delta^{s,n}}\{\hat\theta \in U_{\theta,\epsilon}\}$ as

$$
\begin{aligned}
-\log \mathrm{P}_\theta^{M_\delta^{s,n}}\{\hat\theta \in U_{\theta,\epsilon}\} =& -\log \int \mathrm{P}_\theta^{M_f\times\delta n}(\,d\check\theta)\mathrm{P}_\theta^{L_{\check\theta}\times(1-\delta)n}\{T^n_{\check\theta} \notin U_{\theta,\epsilon}\} \\
\le& -\int \mathrm{P}_\theta^{M_f\times\delta n}(\,d\check\theta)\log\left(\mathrm{P}_\theta^{L_{\check\theta}\times(1-\delta)n}\{T^n_{\check\theta} \notin U_{\theta,\epsilon}\}\right) \\
\le& -\int \mathrm{P}_\theta^{M_f\times\delta n}(\,d\check\theta)\frac{D^{L_{\check\theta}\times(1-\delta)n}(\theta+\xi\epsilon\|\theta)+h(\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n})}{\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n}},
\end{aligned}
$$

where $\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n} := \mathrm{P}^{L_{\check\theta}\times(1-\delta)n}_{\theta+\xi\epsilon,n}\{T_{\check\theta^n} \notin U_{\theta,\epsilon}\}$, and similarly to (34) we can prove the last inequality. For any $\delta_4 > 0$, we have

$$
\begin{aligned}
&\limsup_{n\to\infty} -\frac{1}{n}\log \mathrm{P}_\theta^{\vec M_\delta^s}\{T_n \notin U_{\theta,\epsilon}\} \\
&\le \limsup_{n\to\infty}\int_{\mathbb{R}} \mathrm{P}_\theta^{M_f\times\delta n}(\,d\check\theta)(1-\delta)\min_{\xi=1-\delta_4,-(1-\delta_4)}\frac{(1-\delta)D^{L_{\check\theta}}(\theta+\xi\epsilon\|\theta)+\frac{h(\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n})}{n}}{(1-\delta)\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n}} \\
&= (1-\delta)\min_{\xi=1-\delta_4,-(1-\delta_4)} D^{L_{\check\theta}}(\theta+\xi\epsilon\|\theta) = \frac{1-\delta}{2}J_\theta.
\end{aligned}
$$

The last equation is derived from Lebesgue's convergence theorem and the fact that the probability $\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n}$ tends to 1 uniformly w.r.t. $\check\theta$, as follows from assumptions 1 and 3.

The reason of applicability of Lebesgue's convergence theorem is given as follows. Since $\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n}$ tends to 1 uniformly w.r.t. $\check\theta$, there exists $N, R > 0$ such that $\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n} > \frac{1}{R}, \forall\check\theta \in \Theta, n \ge N$. Thus, we have

$$
\frac{D^{L_{\check\theta}\times(1-\delta)n}(\theta+\xi\epsilon\|\theta)+h(\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n})}{\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n}} \le \frac{R}{1-\delta}\left((1-\delta)D(\theta+\epsilon\xi\|\theta)+2\right) < \infty.
$$

Therefore, we can apply Lebesgue's convergence theorem. Thus, the relations

$$
\begin{aligned}
\alpha(\vec M_\delta^s,\theta) =& \limsup_{\epsilon\to 0}\limsup_{n\to\infty} -\frac{1}{n\epsilon^2}\log \mathrm{P}_\theta^{\vec M_\delta^s}\{T_n \notin U_{\theta,\epsilon}\} \\
&\le (1-\delta)\limsup_{\epsilon\to 0}\frac{1}{\epsilon^2}\min_{\xi=1-\delta_4,-(1-\delta_4)} D^{L_{\check\theta}}(\theta+\xi\epsilon\|\theta) \\
&= (1-\delta)(1-\delta_4)^2\frac{1}{2}J_\theta
\end{aligned}
$$

hold. Since $\delta_4 > 0$ is arbitrary, the inequality (E.6) holds.

Next, we prove the inequality (E.3). Assume that $|\check\theta - \theta| \le \epsilon$ and define

$$
\Lambda(\xi,\check\theta,\theta) := \sup_{\eta\in\mathbb{R}}(\eta\xi - \log\phi_{\theta,\check\theta}(\eta)).
$$

Then, the inequalities

$$
\begin{aligned}
\mathrm{P}_\theta^{L_{\check\theta}\times(1-\delta)n}\{\check\theta \notin U_{\theta,\epsilon}\} \le& \mathrm{P}_\theta^{L_{\check\theta}\times(1-\delta)n}\{|\mathrm{Tr}\,\rho_{\hat\theta}L_{\check\theta} - \mathrm{Tr}\,\rho_\theta L_{\check\theta}| \le (1-\delta_2)\epsilon\} && \text{(E.7)} \\
\le& 2\exp\left(-(1-\delta)n\min\left\{\Lambda((1-\delta_2)\epsilon,\check\theta,\theta),\Lambda(-(1-\delta_2)\epsilon,\check\theta,\theta)\right\}\right) && \text{(E.8)}
\end{aligned}
$$

hold, where (E.7) is derived from fact (iv), and (E.8) is derived from Markov's inequality. Thus,

$$\lim_{n\to\infty} -\frac{1}{n}\log \sup_{\check\theta\in U_{\theta,\sqrt{\epsilon}}} \mathrm{P}_{\theta}^{L_{\check\theta}\times(1-\delta)n}\{\check\theta\notin U_{\theta,\epsilon}\}$$
$$\geq (1-\delta)\inf_{\check\theta\in U_{\theta,\sqrt{\epsilon}}} \min\left\{\Lambda((1-\delta_2)\epsilon,\check\theta,\theta),\Lambda(-(1-\delta_2)\epsilon,\check\theta,\theta)\right\}. \tag{E.9}$$

We let $\epsilon>0$ be a sufficiently small real number for arbitrary $\delta_3>0$ and define $\eta$ by

$$\eta := \epsilon(1-\delta_2)\left(\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\check\theta}}{J_{\check\theta}}-\frac{\mathrm{Tr}\,\rho_\theta L_{\check\theta}}{J_{\check\theta}}\right)^2\right)^{-1}.$$

Then, the inequalities

$$\begin{aligned}
&\Lambda(\pm(1-\delta_2)\epsilon,\check\theta,\theta)\\
\geq\ & \pm(1-\delta_2)\epsilon(\pm\eta)-\log\phi_{\theta,\check\theta}(\pm\eta)\\
\geq\ & \epsilon^2(1-\delta)^2\left(\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\check\theta}}{J_{\check\theta}}-\frac{\mathrm{Tr}\,\rho_\theta L_{\check\theta}}{J_{\check\theta}}\right)^2\right)^{-1}\\
&-\log\left(1+\frac{\epsilon^2(1-\delta)^2}{2}\left(\left(\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\check\theta}}{J_{\check\theta}}-\frac{\mathrm{Tr}\,\rho_\theta L_{\check\theta}}{J_{\check\theta}}\right)^2\right)^{-1}+\delta_3\right)\right)
\end{aligned} \tag{E.10}$$

hold, where (E.10) follows from fact (iii). The uniformity of (E.1) (the fact(iii)) and the boundness of RHS of (E.1) (assumption 3) guarantee that the choice of $\epsilon>0$ is independent of $\theta,\check\theta$. From (E.9) and (E.10), we obtain (E.4) because the function $x\mapsto g(x,y)$ in the case that $y,x\geq 0$.

## Appendix F. Proof of Theorem 13

If the true state is $\rho_{\theta_1}$, the inequalities

$$\begin{aligned}
&\mathrm{P}_{\theta_1}^{M_{\theta_1}^{w,n}}\{T_n\notin U_{\theta_1,\epsilon}\}\\
\leq\ & \mathrm{P}_{\theta_1}^{M_f\times\sqrt{n}}\{\check\theta\notin U_{\theta_1,\epsilon}\}\sup_{\check\theta\notin U_{\theta_1,\epsilon}}\mathrm{P}_{\theta_1}^{E_{\theta_1}^{n-\sqrt{n}}}\left\{e^{n(1-\delta_{n-\sqrt{n}})D(\check\theta\|\theta_1)}\mathrm{P}_{\theta_1}^{E_{\theta_1}^{n-\sqrt{n}}}(\omega)<\mathrm{P}_{\check\theta}^{E_{\theta_1}^{n-\sqrt{n}}}(\omega)\right\}\\
\leq\ & 1\times\sup_{\check\theta\notin U_{\theta_1,\epsilon}} e^{-n(1-\delta_{n-\sqrt{n}})D(\check\theta\|\theta_1)}
\end{aligned}$$

hold. Since $(1-\delta_{n-\sqrt{n}})\to 1$, we have

$$\lim_{n\to\infty}-\frac{1}{n}\log \mathrm{P}_{\theta_1}^{M_{\theta_1}^{w,n}}\{T_n\notin U_{\theta_1,\epsilon}\}=\inf_{\check\theta\notin U_{\theta_1,\epsilon}} D(\check\theta\|\theta_1).$$

Thus, equation (54) is proven. It implies (55).

Next, we show the weak consistency of $\vec{M}_{\theta_1}^{w}$. Assume that the true state $\rho_\theta$ is not $\rho_{\theta_1}$. Then, we have

$$\mathrm{P}_{\theta}^{M_{\theta_1}^{w,n}}\{T_n\notin U_{\theta,\epsilon_n}\}$$

$$\leq \mathrm{P}_\theta^{M_f\times\sqrt{n}}\{\check\theta \notin U_{\theta,\epsilon_n}\}$$
$$+ \mathrm{P}_\theta^{M_f\times\sqrt{n}}\{\check\theta \in U_{\theta,\epsilon_n}\} \sup_{\check\theta\in U_{\theta,\epsilon_n}} \mathrm{P}_\theta^{E_{\theta_1}^{n-\sqrt{n}}} \left\{ e^{n(1-\delta_{n-\sqrt{n}})D(\check\theta\|\theta_1)} \mathrm{P}_{\theta_1}^{E_{\theta_1}^{n-\sqrt{n}}}(\omega) \geq \mathrm{P}_{\check\theta}^{E_{\theta_1}^{n-\sqrt{n}}}(\omega) \right\}, \tag{F.1}$$

where $\epsilon_n := \frac{D(\theta\|\theta_1)}{2\left|\mathrm{Tr}\,\frac{d\rho_\theta}{d\theta}(\log\rho_\theta-\log\rho_{\theta_1})\right|}\delta_n$. Since $\delta_n = \frac{1}{n^{\frac{1}{5}}}$, the convergence $\mathrm{P}_\theta^{M_f\times\sqrt{n}}\{\check\theta \notin U_{\theta,\epsilon_n}\} \to 0$ holds. Also, the relation $U_{\theta,\epsilon_n} \subset U_{\theta,\epsilon_{n-\sqrt{n}}}$ holds. If we can prove

$$\sup_{\check\theta\in U_{\theta,\epsilon_n}} \mathrm{P}_\theta^{E_{\theta_1}^n} \left\{ e^{n(1-\delta_n)D(\check\theta\|\theta_1)} \mathrm{P}_{\theta_1}^{E_{\theta_1}^n}(\omega) \geq \mathrm{P}_{\check\theta}^{E_{\theta_1}^n}(\omega) \right\} \to 0, \tag{F.2}$$

we obtain

$$\mathrm{P}_\theta^{M_{\theta_1}^{w,n}}\{T_n \notin U_{\theta,\epsilon_n}\} \to 0. \tag{F.3}$$

This condition (F.3) is stronger than the weakly consistent condition. Thus, it is sufficient to show (F.2).

From Lemma 14, the relations

$$\begin{aligned}
&\mathrm{P}_\theta^{E_{\theta_1}^n} \left\{ e^{n(1-\delta_n)D(\check\theta\|\theta_1)} \mathrm{P}_{\theta_1}^{E_{\theta_1}^n}(\omega) \geq \mathrm{P}_{\check\theta}^{E_{\theta_1}^n}(\omega) \right\} \\
=& \mathrm{P}_\theta^{E_{\theta_1}^n} \left\{ \frac{1}{n}\left(-\log \mathrm{P}_{\check\theta}^{E_{\theta_1}^n}(\omega) + \log \mathrm{P}_{\theta_1}^{E_{\theta_1}^n}(\omega)\right) + D(\check\theta\|\theta_1) \geq \delta_n D(\check\theta\|\theta_1) \right\} \\
=& \mathrm{P}_\theta^{E_{\theta_1}^n} \Big\{ \frac{1}{n}\left(-\log \mathrm{P}_{\check\theta}^{E_{\theta_1}^n}(\omega) + \log \mathrm{P}_{\theta_1}^{E_{\theta_1}^n}(\omega)\right) + \mathrm{Tr}\,\rho_\theta(\log\rho_{\check\theta} - \log\rho_{\theta_1}) \\
&\quad \geq \delta_n D(\check\theta\|\theta_1) + \mathrm{Tr}(\rho_\theta - \rho_{\check\theta})(\log\rho_{\check\theta} - \log\rho_{\theta_1}) \Big\} \\
\leq& \mathrm{P}_\theta^{E_{\theta_1}^n} \left\{ -\frac{1}{n}\log \mathrm{P}_{\check\theta}^{E_{\theta_1}^n}(\omega) + \mathrm{Tr}\,\rho_\theta\log\rho_{\check\theta} \geq \delta_n D(\check\theta\|\theta_1) + \mathrm{Tr}(\rho_\theta - \rho_{\check\theta})(\log\rho_{\check\theta} - \log\rho_{\theta_1}) \right\} \\
&+ \mathrm{P}_\theta^{E_{\theta_1}^n} \left\{ \frac{1}{n}\log \mathrm{P}_{\theta_1}^{E_{\theta_1}^n}(\omega) - \mathrm{Tr}\,\rho_\theta\log\rho_{\theta_1} \geq \delta_n D(\check\theta\|\theta_1) + \mathrm{Tr}(\rho_\theta - \rho_{\check\theta})(\log\rho_{\check\theta} - \log\rho_{\theta_1}) \right\} \\
\leq& \exp - \Big( n \sup_{0\leq t\leq 1} \left(\delta_n D(\check\theta\|\theta_1) + \mathrm{Tr}(\rho_\theta - \rho_{\check\theta})(\log\rho_{\check\theta} - \log\rho_{\theta_1}) - \mathrm{Tr}\,\rho_\theta\log\rho_{\check\theta}\right) t \\
&\quad - t\frac{(k+1)\log(n+1)}{n} - \log\mathrm{Tr}\,\rho_\theta\rho_{\check\theta}^{-t} \Big) \\
&+ \exp - \left( n \sup_{0\leq t} \left(\delta_n D(\check\theta\|\theta_1) + \mathrm{Tr}(\rho_\theta - \rho_{\check\theta})(\log\rho_{\check\theta} - \log\rho_{\theta_1}) + \mathrm{Tr}\,\rho_\theta\log\rho_{\theta_1}\right) t - \log\mathrm{Tr}\,\rho_\theta\rho_{\theta_1}^t \right)
\end{aligned} \tag{F.4}$$

hold. In the following, we assume that $|\theta - \check\theta| \leq \epsilon_n$. Since $\epsilon_n = \frac{D(\theta\|\theta_1)}{2\left|\mathrm{Tr}\,\frac{d\rho_\theta}{d\theta}(\log\rho_\theta-\log\rho_{\theta_1})\right|}\delta_n$, we can derive that $\delta_n D(\check\theta\|\theta_1) + \mathrm{Tr}(\rho_\theta - \rho_{\check\theta})(\log\rho_{\check\theta} - \log\rho_{\theta_1}) \leq \frac{1}{2}D(\theta\|\theta_1)\delta_n + O(\delta_n^2)$. Substituting $t = s\delta_n$, we have

$$\begin{aligned}
&\sup_{\check\theta\in U_{\theta,\epsilon_n}} \frac{1}{n\delta_n^2}\Big( n \sup_{0\leq t\leq 1} (\delta_n D(\check\theta\|\theta_1) + \mathrm{Tr}(\rho_\theta - \rho_{\check\theta})(\log\rho_{\check\theta} - \log\rho_{\theta_1}) - \mathrm{Tr}\,\rho_\theta\log\rho_{\check\theta})t \\
&- t\frac{(k+1)\log(n+1)}{n} - \log\mathrm{Tr}\,\rho_\theta\rho_{\check\theta}^{-t} \Big)
\end{aligned}$$

$$
\begin{aligned}
\geq & \sup_{\check{\theta}\in U_{\theta,\epsilon_n}} \frac{1}{\delta_n^2}\Bigg( (\frac{1}{2}D(\theta\|\theta_1)\delta_n + O(\delta_n^2) - \operatorname{Tr}\rho_\theta \log\rho_{\check{\theta}}) s\delta_n - s\delta_n \frac{(k+1)\log(n+1)}{n} \\
& + \operatorname{Tr}\rho_\theta\log\rho_{\check{\theta}} s\delta_n - \frac{1}{2}(\operatorname{Tr}\rho_\theta(\log\rho_{\check{\theta}})^2 - (\operatorname{Tr}\rho_\theta\log\rho_{\check{\theta}})^2)s^2\delta_n^2 + O(\delta_n^3)\Bigg) \\
\geq & \sup_{\check{\theta}\in U_{\theta,\epsilon_n}} \frac{1}{\delta_n^2}\Bigg( \frac{1}{2}D(\theta\|\theta_1)s\delta_n^2 + O(\delta_n^3) - s\delta_n \frac{(k+1)\log(n+1)}{n} \\
& - \frac{1}{2}(\operatorname{Tr}\rho_\theta(\log\rho_{\check{\theta}})^2 - (\operatorname{Tr}\rho_\theta\log\rho_{\check{\theta}})^2)s^2\delta_n^2 + O(\delta_n^3)\Bigg) \\
\to & \frac{1}{2}D(\theta\|\theta_1)s - \frac{1}{2}\left(\operatorname{Tr}\rho_\theta(\log\rho_\theta)^2 - (\operatorname{Tr}\rho_\theta\log\rho_\theta)^2\right)s^2 \\
= & -\frac{1}{2}\left(\operatorname{Tr}\rho_\theta(\log\rho_\theta)^2 - (\operatorname{Tr}\rho_\theta\log\rho_\theta)^2\right)\left(s - \frac{D(\theta\|\theta_1)}{2(\operatorname{Tr}\rho_\theta(\log\rho_\theta)^2 - (\operatorname{Tr}\rho_\theta\log\rho_\theta)^2)}\right)^2 \\
& + \frac{D(\theta\|\theta_1)^2}{8(\operatorname{Tr}\rho_\theta(\log\rho_\theta)^2 - (\operatorname{Tr}\rho_\theta\log\rho_\theta)^2)}.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
& \lim_{n\to\infty}\sup_{\check{\theta}\in U_{\theta,\epsilon_n}} \frac{1}{n\delta_n^2}\Bigg( n\sup_{0\leq t\leq 1}(\delta_n D(\check{\theta}\|\theta_1) + \operatorname{Tr}(\rho_\theta - \rho_{\check{\theta}})(\log\rho_{\check{\theta}} - \log\rho_{\theta_1}) - \operatorname{Tr}\rho_\theta\log\rho_{\check{\theta}})t \\
& - t\frac{(k+1)\log(n+1)}{n} - \log\operatorname{Tr}\rho_\theta\rho_{\check{\theta}}^{-t}\Bigg) \\
\geq & \frac{D(\theta\|\theta_1)^2}{8(\operatorname{Tr}\rho_\theta(\log\rho_\theta)^2 - (\operatorname{Tr}\rho_\theta\log\rho_\theta)^2)} > 0.
\end{aligned} \tag{F.5}
$$

Also, we obtain

$$
\begin{aligned}
& \sup_{\check{\theta}\in U_{\theta,\epsilon_n}} \frac{1}{n\delta_n^2}\left( n\sup_{0\leq t}(\delta_n D(\check{\theta}\|\theta_1) + \operatorname{Tr}(\rho_\theta - \rho_{\check{\theta}})(\log\rho_{\check{\theta}} - \log\rho_{\theta_1}) + \operatorname{Tr}\rho_\theta\log\rho_{\theta_1})t - \log\operatorname{Tr}\rho_\theta\rho_{\theta_1}^t\right) \\
\geq & \sup_{\check{\theta}\in U_{\theta,\epsilon_n}} \frac{1}{\delta_n^2}\Bigg( (\frac{1}{2}D(\theta\|\theta_1)\delta_n + O(\delta_n^2) + \operatorname{Tr}\rho_\theta\log\rho_{\theta_1})s\delta_n - \operatorname{Tr}\rho_\theta\log\rho_{\theta_1}s\delta_n \\
& - \frac{1}{2}(\operatorname{Tr}\rho_\theta(\log\rho_{\theta_1})^2 - (\operatorname{Tr}\rho_\theta\log\rho_{\theta_1})^2)s^2\delta_n^2 + O(\delta_n^3)\Bigg) \\
= & \sup_{\check{\theta}\in U_{\theta,\epsilon_n}} \frac{1}{\delta_n^2}\left(\left(\frac{1}{2}D(\theta\|\theta_1)s - \frac{1}{2}\left(\operatorname{Tr}\rho_\theta(\log\rho_{\theta_1})^2 - (\operatorname{Tr}\rho_\theta\log\rho_{\theta_1})^2\right)s^2\right)\delta_n^2 + O(\delta_n^3)\right) \\
\to & \frac{1}{2}D(\theta\|\theta_1)s - \frac{1}{2}(\operatorname{Tr}\rho_\theta(\log\rho_{\theta_1})^2 - (\operatorname{Tr}\rho_\theta\log\rho_{\theta_1})^2)s^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& \lim_{n\to\infty}\sup_{\check{\theta}\in U_{\theta,\epsilon_n}} \frac{1}{n\delta_n^2}\left( n\sup_{0\leq t}(\delta_n D(\check{\theta}\|\theta_1) + \operatorname{Tr}(\rho_\theta - \rho_{\check{\theta}})(\log\rho_{\check{\theta}} - \log\rho_{\theta_1}) + \operatorname{Tr}\rho_\theta\log\rho_{\theta_1})t - \log\operatorname{Tr}\rho_\theta\rho_{\theta_1}^t\right) \\
\geq & \frac{D(\theta\|\theta_1)^2}{8(\operatorname{Tr}\rho_\theta(\log\rho_{\theta_1})^2 - (\operatorname{Tr}\rho_\theta\log\rho_{\theta_1})^2)} > 0.
\end{aligned} \tag{F.6}
$$

Since $n\delta_n^2 \to \infty$, the relation (F.2) follows from (F.4),(F.5) and (F.6).

## Appendix G. Pinching map and group theoretical viewpoint

### *Appendix G.1. Pinching map in non-asymptotic setting*

In the following, we prove Lemma 14 and construct the PVM $E_\theta^n$ after some discussions about the pinching map in the non-asymptotic setting and group representation theory. In this subsection, we present some definitions and non-asymptotic discussions.

A state $\rho$ is called *commutative* with a PVM $E(=\{E_i\})$ on $\mathcal{H}$ if $\rho E_i = E_i \rho$ for any index $i$. For PVMs $E(=\{E_i\}_{i\in I}), F(=\{F_j\}_{j\in J})$, the notation $E \le F$ means that for any index $i \in I$ there exists a subset $(F/E)_i$ of the index set $J$ such that $E_i = \sum_{j\in(F/E)_i} F_j$. For a state $\rho$, we denote by $E(\rho)$ the spectral measure of $\rho$ which can be regarded as a PVM. The pinching map $\mathcal{E}_E$ with respect to a PVM $E$ is defined as:

$$\mathcal{E}_E : \rho \mapsto \sum_i E_i \rho E_i, \tag{G.1}$$

which is an affine map from the set of states to itself. Note that the state $\mathcal{E}_E(\rho)$ is commutative with a PVM $E$. If a PVM $F = \{F_j\}_{j\in J}$ is commutative with a PVM $E = \{E_i\}_{i\in I}$, we can define the PVM $F \times E = \{F_j E_i\}_{(i,j)\in I\times J}$, which satisfies that $F \times E \ge E$ and $F \times E \ge F$. For any PVM $E$, the supremum of the dimension of $E_i$ is denoted by $w(E)$.

**Lemma 22** *Let $E$ be a PVM such that $w(E) < \infty$. If states $\sigma$ and $\rho$ are commutative with the PVM $E$, and if a PVM $F$ satisfies $E \le F, E(\sigma) \le F$, then we have*

$$D(\rho\|\sigma) - \log w(E) \le D(\mathcal{E}_F(\rho)\|\mathcal{E}_F(\sigma)) \le D(\rho\|\sigma).$$

This lemma follows from Lemma 23 and Lemma 24 below.

**Lemma 23** *Let $\rho, \sigma$ be states. If a PVM $F$ satisfies $E(\sigma) \le F$, then*

$$D(\rho\|\sigma) = D(\mathcal{E}_F(\rho)\|\mathcal{E}_F(\sigma)) + D(\rho\|\mathcal{E}_F(\rho)). \tag{G.2}$$

*Proof:* Since $E(\sigma) \le F$ and $F$ is commutative with $\sigma$, we have $\operatorname{Tr}\mathcal{E}_F(\rho)\log\mathcal{E}_F(\sigma) = \operatorname{Tr}\rho\log\sigma$. Since $\rho$ is commutative with $\log\rho$, we have $\operatorname{Tr}\mathcal{E}_F(\rho)\log\rho = \operatorname{Tr}\rho\log\rho$. Therefore, we obtain the following:

$$\begin{aligned} D(\mathcal{E}_F(\rho)\|\mathcal{E}_F(\sigma)) - D(\rho\|\sigma) &= \operatorname{Tr}\mathcal{E}_F(\rho)(\log\mathcal{E}_F(\rho) - \log\mathcal{E}_F(\sigma)) - \operatorname{Tr}\rho(\log\rho - \log\sigma) \\ &= \operatorname{Tr}\mathcal{E}_F(\rho)(\log\mathcal{E}_F(\rho) - \log\rho). \end{aligned}$$

This proves (G.2). ■

**Lemma 24** *Let $E, F$ be PVMs such that $E \le F$. If a state $\rho$ is commutative with $E$, we have*

$$D(\rho\|\mathcal{E}_F(\rho)) \le \log w(E). \tag{G.3}$$

*Proof:* Let $a_i := \operatorname{Tr} E_i \rho E_i, \rho_i := \frac{1}{a_i} E_i \rho E_i$. Then, we have $\rho = \sum_i a_i \rho_i$, $\mathcal{E}_F(\rho) = \sum_i a_i \mathcal{E}_F(\rho_i)$, $\sum_i a_i = 1$. Therefore,

$$D(\rho\|\mathcal{E}_F(\rho)) = \sum_i \operatorname{Tr} E_i \rho(\log\rho - \log\mathcal{E}_F(\rho)) = \sum_i \operatorname{Tr} E_i\rho E_i(E_i\log\rho E_i - E_i\log\mathcal{E}_F(\rho)E_i)$$

$$= \sum_i a_i D(\rho_i \| \mathcal{E}_F(\rho_i)) \le \sup_i D(\rho_i \| \mathcal{E}_F(\rho_i)) = \sup_i (\operatorname{Tr} \rho_i \log \rho_i - \operatorname{Tr} \mathcal{E}_F(\rho_i) \log \mathcal{E}_F(\rho_i))$$

$$\le -\sup_i \operatorname{Tr} \mathcal{E}_F(\rho_i) \log \mathcal{E}_F(\rho_i) \le \sup_i \log \dim E_i = \log w(E).$$

Thus, we obtain inequality (G.3). ■

Let us consider another type of inequality.

**Lemma 25** *Let $E$ be a PVM such that $w(E) < \infty$. If the state $\rho$ is commutative with $E$, and if a PVM $M$ satisfies that $M \ge E$, we have*

$$\rho \le \mathcal{E}_M(\rho) w(E) \tag{G.4}$$

$$\rho^{-t} \ge \mathcal{E}_M(\rho)^{-t} w(E)^{-t} \tag{G.5}$$

*for* $1 \le t \le 0$.

*Proof:* It is sufficient for (G.4) to show

$$\rho \le k \mathcal{E}_M(\rho), \tag{G.6}$$

for any state $\rho$ and any PVM $M$ on a $k$-dimensional Hilbert space $\mathcal{H}$. Now, it is sufficient to prove (G.6) in the pure state case. For any $\phi, \psi \in \mathcal{H}$, we have

$$\langle \psi | k \mathcal{E}_M(|\phi\rangle\langle\phi|) - |\phi\rangle\langle\phi| |\psi\rangle = k \sum_{i=1}^{k} \langle \psi | M_i | \phi \rangle \langle \phi | M_i | \psi \rangle - \left| \sum_{i=1}^{k} \langle \psi | M_i | \phi \rangle \right|^2 \ge 0.$$

The last inequality follows from Schwartz inequality for vectors $\{\langle \psi | M_i | \phi \rangle\}_{i=1}^{k}$ and $\{1\}_{i=1}^{k}$. It is well known that the function $u \mapsto -u^{-t}$ $(0 \le t \le 1)$ is an operator monotone function [32]. Thus, (G.4) implies (G.5). ■

**Lemma 26** *If a PVM $M$ is commutative with a state $\sigma$ and $w(M) = 1$, we have*

$$\mathrm{P}_{\rho}^{M} \left\{ \log \mathrm{P}_{\sigma}^{M}(\omega) \ge a \right\} \le \exp \left( -\sup_{0 \le t} \left( at - \log \operatorname{Tr} \rho \sigma^t \right) \right) \tag{G.7}$$

*for any state* $\rho$.

*Proof:* From Markov's inequality, we have

$$p\{X \ge a\} \le \exp -\Lambda_t(X, p, a) \tag{G.8}$$

$$\Lambda_t(X, p, a) := at - \log \int e^{tX(\omega)} p(\, d\omega).$$

Since $w(M) = 1$, the relation $\sum_\omega \mathrm{P}_\rho^M(\omega) \mathrm{P}_\sigma^M(\omega)^t = \operatorname{Tr} \mathcal{E}_M(\rho) \mathcal{E}_M(\sigma)^t$ holds. It yields

$$\Lambda_t(\log \mathrm{P}_\sigma^M, \mathrm{P}_\rho^M, a) = at - \log \operatorname{Tr} \mathcal{E}_M(\rho) \mathcal{E}_M(\sigma)^t = at - \log \operatorname{Tr} \rho \sigma^t.$$

Thus, we obtain (G.7). ■

**Lemma 27** *Assume that $E$ and $M$ are PVMs such that $w(E) < \infty, w(M) = 1$ and $M \ge E$. If the states $\rho$ and $\rho'$ are commutative with $E$, we have*

$$\mathrm{P}_{\rho}^{M} \left\{ -\log \mathrm{P}_{\rho'}^{M}(\omega) \ge a \right\} \le \exp \left( -\sup_{0 \le t \le 1} \left( (a - \log w(E)) t - \log \operatorname{Tr} \rho \rho'^{-t} \right) \right). \tag{G.9}$$

*Proof:* If $0 \le t \le 1$, we have

$$\begin{aligned}\Lambda_t(-\log \mathrm{P}^M_{\rho'}, \mathrm{P}^M_{\rho}, a) &= at - \log \operatorname{Tr} \mathcal{E}_M(\rho)\mathcal{E}_M(\rho')^{-t} = at - \log \operatorname{Tr} \rho \mathcal{E}_M(\rho')^{-t} \\ &\ge at - \log w(E)^t \operatorname{Tr} \rho \rho'^{-t} \qquad \text{(G.10)} \\ &\ge (a - \log w(E))t - \log \operatorname{Tr} \rho\rho'^{-t}, \qquad \text{(G.11)}\end{aligned}$$

where (G.10) follows from Lemma 25. Therefore, from (G.8) and (G.11), we obtain (G.9). ■

### *Appendix G.2. group representation and its irreducible decomposition*

In this subsection, we consider the relation between irreducible representations and PVMs for the purpose of constructing the PVM $E^n_\theta$ and a proof of Lemma 14. Let $V$ be a finite dimensional vector space over the complex numbers $\mathbb{C}$. A map $\pi$ from a group $G$ to the generalized linear group of a vector space $V$ is called a *representation* on $V$ if the map $\pi$ is homomorphic, i.e., $\pi(g_1)\pi(g_2) = \pi(g_1 g_2)$, $\forall g_1, g_2 \in G$. The subspace $W$ of $V$ is called *invariant* with respect to a representation $\pi$ if the vector $\pi(g)w$ belongs to the subspace $W$ for any vector $w \in W$ and any element $g \in G$. The representation $\pi$ is called *irreducible* if there is no proper nonzero invariant subspace of $V$ with respect to $\pi$. Let $\pi_1$ and $\pi_2$ be representations of a group $G$ on $V_1$ and $V_2$, respectively. The *tensored* representation $\pi_1 \otimes \pi_2$ of $G$ on $V_1 \otimes V_2$ is defined as $(\pi_1 \otimes \pi_2)(g) = \pi_1(g) \otimes \pi_2(g)$, and the *direct sum* representation $\pi_1 \oplus \pi_2$ of $G$ on $V_1 \oplus V_2$ is also defined as $(\pi_1 \oplus \pi_2)(g) = \pi_1(g) \oplus \pi_2(g)$.

In the following, we treat a representation $\pi$ of a group $G$ on a finite-dimensional Hilbert space $\mathcal{H}$. The following fact is crucial in later arguments. There exists an irreducible decomposition $\mathcal{H} = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_l$ such that the irreducible components are orthogonal to one another if for any element $g \in G$ there exists an element $g^* \in G$ such that $\pi(g)^* = \pi(g^*)$, where $\pi(g)^*$ denotes the adjoint of the linear map $\pi(g)$. We can regard the irreducible decomposition $\mathcal{H} = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_l$ as the PVM $\{P_{\mathcal{H}_i}\}_{i=1}^l$, where $P_{\mathcal{H}_i}$ denotes the projection to $\mathcal{H}_i$. If two representations $\pi_1, \pi_2$ satisfy the preceding condition, the tensored representation $\pi_1 \otimes \pi_2$, also satisfies it. Note that in general, an irreducible decomposition of a representation satisfying the preceding condition is not unique. In other words, we cannot uniquely define the PVM from such a representation.

### *Appendix G.3. Construction of PVM $E^n_\theta$ and the tensored representation*

In this subsection, we construct the PVM $E^n_\theta$ after the discussion of the tensored representation. Let the dimension of the Hilbert space $\mathcal{H}$ be $k$. Concerning the natural representation $\pi_{\mathrm{SL}(\mathcal{H})}$ of the special linear group $\mathrm{SL}(\mathcal{H})$ on $\mathcal{H}$, we consider its $n$-th tensored representation $\pi^{\otimes n}_{\mathrm{SL}(\mathcal{H})} := \underbrace{\pi_{\mathrm{SL}(\mathcal{H})} \otimes \cdots \otimes \pi_{\mathrm{SL}(\mathcal{H})}}_{n}$ on the tensored space $\mathcal{H}^{\otimes n}$. For any element $g \in \mathrm{SL}(\mathcal{H})$, the relation $\pi_{\mathrm{SL}(\mathcal{H})}(g)^* = \pi_{\mathrm{SL}(\mathcal{H})}(g^*)$ holds where the element $g^* \in \mathrm{SL}(\mathcal{H})$ denotes the adjoint matrix of the matrix $g$. Consequently, there exists an irreducible decomposition of $\pi_{\mathrm{SL}(\mathcal{H})}^{\otimes n}$ regarded as a PVM and we denote by $Ir^{\otimes n}$ the set of such PVMs.

From Weyl's dimension formula ((7.1.8) or (7.1.17) in Weyl [33] Goodman-Wallch [34]), the $n$-th symmetric tensored space is the maximum-dimensional space in the irreducible subspaces with respect to the $n$-th tensored representation $\pi_{\mathrm{SL}(\mathcal{H})}^{\otimes n}$. Its dimension equals the repeated combination ${}_kH_n$ evaluated by ${}_kH_n = \binom{n+k-1}{k-1} = \binom{n+k-1}{n} = {}_{n+1}H_{k-1} \le (n+1)^{k-1}$. Thus, any element $E^n \in Ir^{\otimes n}$ satisfies the following:

$$w(E^n) \le (n+1)^{k-1}. \tag{G.12}$$

**Lemma 28** *A PVM $E^n \in Ir^{\otimes n}$ is commutative with the $n$-th tensored state $\rho^{\otimes n}$ of any state $\rho$ on $\mathcal{H}$.*

*Proof:* If $\det \rho \neq 0$, this lemma is trivial based on the fact that $\det(\rho)^{-1}\rho \in \mathrm{SL}(\mathcal{H})$. If $\det \rho = 0$, there exists a sequence $\{\rho_i\}_{i=1}^{\infty}$ such that $\det \rho_i \neq 0$ and $\rho_i \to \rho$ as $i \to \infty$. We have $\rho_i^{\otimes n} \to \rho^{\otimes n}$ as $i \to \infty$. Because a PVM $E^n \in Ir^{\otimes n}$ is commutative with $\rho_i^{\otimes n}$, it is also commutative with $\rho^{\otimes n}$. ∎

**Definition 29** *We can define the PVM $E^n \times E(\rho^{\otimes n})$ for any PVM $E^n \in Ir^{\otimes n}$. Now we define the PVM $E_\theta^n$ satisfying $w(E_\theta^n) = 1$, $E_\theta^n \ge E^n \times E(\rho_\theta^{\otimes n})$ for a PVM $E^n \in Ir^{\otimes n}$. Note that the $E_\theta^n$ is not unique.*

*Proof of Lemma 14:* From Lemmas 26, 27, (G.12) and the definition of $E_\theta^n$, we obtain Lemma 14. ∎

*Proof of Lemma 19:* From Lemma 22, (G.12) and the definition of $E_\theta^n$, we obtain Lemma 19. ∎

## Appendix H. Large deviation theory for exponential family

In this section, we summarize the large deviation theory for an exponential family. A $d$-dimensional probability family is called an exponential family if there exist linearly independent real-valued random variables $F_1, \ldots, F_d$ and a probability distribution $p$ on the probability space $\Omega$ such that the family consists of the probability distribution

$$p_\theta(\,d\omega) := \exp\left(\sum_{i=1}^{d} \theta^i F_i(\omega) - \psi(\theta)\right) p(\,d\omega)$$

$$\psi(\theta) := \log \int_\Omega \exp\left(\sum_{i=1}^{d} \theta^i F_i(\omega)\right) p(\,d\omega).$$

In this family, the parametric space is given by $\Theta := \{\theta \in \mathbb{R}^d | 0, < \psi(\theta) < \infty\}$, the parameter $\theta$ is called the natural parameter and the function $\psi(\theta)$ is called the potential. We define the dual potential $\phi(\theta)$ and the dual parameter $\eta(\theta)$, called the expectation parameter, as

$$\eta_i(\theta) := \frac{\partial \psi(\theta)}{\partial \theta^i} = \log \int_\Omega F_i(\omega) p_\theta(\,d\omega)$$

$$\phi(\theta) := \max_{\theta'} \sum_{i=1}^{d} \theta'^i \eta_i(\theta) - \psi(\theta').$$

From (H.1), we have

$$\phi(\theta) = \sum_{i=1}^{d} \theta^i \eta_i(\theta) - \psi(\theta).$$

In this family, the sufficient statistics are given by $F_1(\omega), \ldots, F_d(\omega)$. The MLE $\hat{\theta}(\omega)$ is given by $\eta_i(\hat{\theta}(\omega)) = F_i(\omega)$. The KL divergence $D(\theta\|\theta_0) := D(p_\theta\|p_{\theta_0})$ is calculated by

$$\begin{aligned}
D(\theta\|\theta_0) = & \int_\Omega \log \frac{p_\theta(\omega)}{p_{\theta_0}(\omega)} p_\theta(\, d\omega) = \int_\Omega \sum_i (\theta^i - \theta_0^i) F_i(\omega) + \psi(\theta_0) - \psi(\theta) p_\theta(\, d\omega) \\
= & \sum_i (\theta^i - \theta_0^i)\eta_i(\omega) + \psi(\theta_0) - \psi(\theta) = \phi(\theta) + \psi(\theta_0) - \sum_i \theta_0^i \eta_i(\omega) \\
= & \max_{\theta'} \sum_i \theta'^i \eta_i(\theta) - \psi(\theta') + \psi(\theta_0) - \sum_i \theta_0^i \eta_i(\theta) \\
= & \max_{\theta'} \sum_i (\theta'^i - \theta_0'^i)\eta_i(\theta) - \log \int_\Omega \exp\left( \sum_i (\theta^i - \theta_0^i) F_i(\omega) \right) p_\theta(\, d\omega).
\end{aligned}$$

Next, we discuss the $n$-i.i.d. extension of the family $\{p_\theta | \theta \in \Theta\}$. For the data $\vec{\omega}_n := (\omega_1, \ldots, \omega_n) \in \Omega^n$, the probability distribution $p_\theta^n(\vec{\omega}_n) := p_\theta(\omega_1) \ldots p_\theta(\omega_n)$ is written by

$$\begin{aligned}
& p_\theta^n(\vec{\omega}_n) = \exp\left( n \sum_i \theta^i F_{n,i}(\vec{\omega}_n) - n\psi(\theta) \right) p^n(\, d\vec{\omega}_n) \\
& p^n(\, d\vec{\omega}_n) := p(\, d\omega_1) \ldots p(\, d\omega_n) \\
& F_{n,i}(\vec{\omega}_n) := \frac{1}{n} \sum_{k=1}^{n} F_i(\omega_k).
\end{aligned}$$

Since the expectation parameter of the probability family $\{p_\theta^n | \theta \in \Theta\}$ is given by $n\eta_i(\theta)$, the MLE $\hat{\theta}_n(\vec{\omega}_n)$ is given by

$$n\eta_i(\hat{\theta}_n(\vec{\omega}_n)) = nF_{n,i}(\vec{\omega}_n). \tag{H.1}$$

Applying Cramér's Theorem [29] to the random variables $F_1, \ldots, F_d$ and the distribution $p_{\theta_0}$, for any subset $S \subset \mathbb{R}^d$ we have

$$\begin{aligned}
\inf_{\eta \in \overline{S}} \sup_{\theta' \in \mathbb{R}^d} \sum_i \theta'^i (\eta_i - \mathrm{E}_{\theta_0}(F_i)) - \psi_{\theta_0}(\theta') \le & \lim_{n\to\infty} \frac{-1}{n} \log p_{\theta_0}^n \{\vec{F}_n \in S\} \\
\le & \inf_{\eta \in \mathrm{int} S} \sup_{\theta' \in \mathbb{R}^d} \sum_i \theta'^i (\eta_i - \mathrm{E}_{\theta_0}(F_i)) - \psi_{\theta_0}(\theta'),
\end{aligned}$$

where

$$\begin{aligned}
& \mathrm{E}_{\theta_0}(F_i)) := \int_\Omega F_i(\omega) p_\theta(\, d\omega) \\
& \psi_{\theta_0}(\theta) := \int_\Omega \exp\left( \sum_i \theta^i F_i(\omega) \right) p_\theta(\, d\omega) \\
& \vec{F}_n(\vec{\omega}_n) := (F_{n,1}(\vec{\omega}_n), \ldots, F_{n,d}(\vec{\omega}_n)),
\end{aligned}$$

and int$S$ denotes the interior of $S$, which is consistent with $(\overline{S^c})^c$. Since

$$\sup_{\theta' \in \mathbb{R}^d} \sum_i \theta'^i (\eta_i - \mathrm{E}_{\theta_0}(F_i)) - \psi_{\theta_0}(\theta') = \sup_{\theta' \in \mathbb{R}^d} \sum_i \theta'^i (\eta_i - \eta_i(\theta_0)) - \psi(\theta') + \psi(\theta_0) = D(\theta \| \theta_0)$$

and the map $\theta \mapsto D(\theta\|\theta_0)$ is continuous, it follows from (H.1) that

$$\lim_{n\to\infty} \frac{-1}{n} \log p^n_{\theta_0}\{\hat{\theta}_n \in \Theta'\} = \inf_{\theta\in\Theta'} D(\theta\|\theta_0)$$

for any subset $\Theta' \subset \Theta$, which is equivalent with (74). Conversely, if an estimator $\{T_n(\vec{\omega}_n)\}$ satisfies the weak consistency

$$\lim_{n\to\infty} p^n_\theta\{\|T_n(\vec{\omega}_n) - \theta\| > \epsilon\} \to 0, \quad \forall\epsilon > 0, \forall\theta \in \Theta,$$

then, similarly to (32), we can prove

$$\lim_{n\to\infty} \frac{-1}{n} \log p^n_{\theta_0}\{T_n(\vec{\omega}_n) \in \Theta'\} \le \inf_{\theta\in\Theta'} D(\theta\|\theta_0).$$

Therefore, we can conclude that the MLE is optimal in the large deviation sense for exponential families.

## References


[1] Holevo A S 1998 *IEEE Trans.* **IT-44**, 269, (Originally appeared in LANL eprint quant-ph/9611023).

[2] Schumacher B and Westmoreland M D 1997 *Phys. Rev.* A, **56**, 131.

[3] Schumacher B 1995 *Phys. Rev.* A, **51**, 2738.

[4] Winter A 2000 "Coding Theorems of Quantum Information Theory," Ph.D. dissertation, Uni Bielefeld, (Originally appeared in quant-ph/9907077).

[5] Hayashi M 2002 "Exponents of quantum fixed-length pure state source coding," LANL eprint quant-ph/0202??? (2002).

[6] Hiai F and Petz D 1991 *Commun. Math. Phys.* **143**, 99-114.

[7] Ogawa T and Nagaoka H 2000 *IEEE Trans.* **IT-46**, 2428, (Originally appeared in LANL eprint quant-ph/9808063).

[8] Helstrom C W 1967 *Phys. Lett.* **25A**, 101.

[9] Helstrom C W 1976 *Quantum Detection and Estimation Theory,* (Academic Press, New York).

[10] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory*, (North_Holland, Amsterdam).

[11] Gill R and Massar S 2000 *Phys. Rev.* A, **61**, 042312.

[12] Hayashi M 1997 "A Linear Programming Approach to Attainable Cramer-Rao type bound" (in *Quantum Communication, Computing, and Measurement*, edited by O. Hirota, A. S. Holevo, and C. M. Caves, Plenum Publishing pp. 99-108.).

[13] Hayashi M 1997 LANL eprint quant-ph/9704044.

[14] Nagaoka H 1988 "An Asymptotically Efficient Estimator for a One-dimensional Parametric Model of Quantum Statistical Operators," Proc. of 1988 IEEE International Symposium on Information Theory, p.198.

[15] Bahadur R R 1971 *Some limit theorems in statistics,* Regional Conf. Series in Applied Mathematics, no 4, SIAM, Philadelphia.

[16] Fu J C 1973 *Ann. Stat.* **1**, 745.

[17] Petz D 1996 *Linear Algebra and Its Applications*, **224**, 81.

[18] Amari S and Nagaoka H 2000 *Methods of Information Geometry,* AMS & Oxford University Press.

[19] Petz D and Toth G 1993 *Lett. Math. Phys.* **27**, 205.

[20] Petz D and Sudár C 1998 "Extending the Fisher metric to density matrices," In O. E. Barndorff-Nielsen and E. B. V. Jensen (eds.), *Geometry in Present Day Science*, 21, Singapore: World Scientific.
[21] Hayashi M 2001 *J. Phys. A: Math. and Gen.* **34**, 3413, (Originally appeared in LANL eprint quant-ph/9704040).
[22] Lindblad G 1975 *Commun. Math. Phys.* **40**, 147.
[23] Uhlmann A 1977 *Commun. Math. Phys.* **54**, 21.
[24] Uhlmann A 1993 *Rep. Math. Phys.* **33**, 253-263.
[25] Matsumoto K 1995 "Geometry of quantum states," Master's thesis, Tokyo University. (In Japanese)
[26] Akahira M and Takeuchi K 1991 *Journal of Computing and Information*, **2** 71.
[27] Jozsa R 1994 *J. Mod. Opt.* **41**, 2315.
[28] Fuchs C A 1995 "Distinguishability and accessible information in quantum theory," Ph. D. Dissertation, University of New Mexico. LANL eprint quant-ph/9601020.
[29] Backlew J A 1990 *Large Deviation Techniques in Decision, Simulation and Estimation,* John Wiley & Sons, New York.
[30] Hayashi M and Matsumoto K 2000 *IEICE* **J83-A**, No.6, 629. (In Japanese)
[31] Barndorff-Nielsen O E 1978 *Information and Exponential Families in Statistical Theory,* Wiley, New York.
[32] Bhatia R 1997 *Matrix Analysis*, Springer, New York.
[33] Weyl H 1939 *The classical groups, their invariants and representations,* Princeton.
[34] Goodman R and Wallch N 1998 *Representations and Invariants of the Classical Groups,* Cambridge University Press.